ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Architectural and Landscape Heritage (33th Cycle)

Photogrammetry and Deep Learning to improve Cultural Heritage records: extracting 3D metric information from historical images

Francesca Condorelli

* * * * * *

Supervisors
Prof. Fulvio Rinaudo, Supervisor
Prof. Rosa Tamborrino, Co-Supervisor

Politecnico di Torino
May 25, 2021

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

…………………………………......
Francesca Condorelli
Turin, May 25, 2021

Summary

The thesis aims to give an analysis and assessment through the extraction of metric information from historical images and to experiment with its potentialities in the heritage field with the aim of valorising historical iconographical documentation.

This thesis deals with historical images stored in historical archives in the heritage context. This documentation has been produced for many other purposes but actually contains something very valuable for Cultural Heritage: data and information. Specifically, the thesis explores how to extract and use dimensional data from historical images for documenting monuments, buildings and groups of buildings that no longer exist or were transformed over time.

This thesis focuses on two kinds of images documentation, historical photographs and film footage from the early 19th century. Concerning the extraction of metric information from historical images, this thesis intends to give an upgrade of the previous studies on this topic using the latest developed technologies. This upgrade is in terms of metric precision provided with the use of classical photogrammetry combined with Deep Learning solutions. The output is a method, suitable for different fields, but experimented with as an application in the heritage context.

In this proposed method Deep Learning is used for the retrieval of primary data used as input material in the standard Structure-from-Motion (SfM) pipeline used to reconstruct lost Cultural Heritage. Object detection Neural Networks were trained to automatically recognise a specific monument in film footage and image collections. Then the images suitable to be processed with photogrammetry are selected from all the frames detected by the Neural Network. The selection is performed according to the camera motions within the scene of the video. Only the shots taken from multiple points of view of the same scene are suitable for the photogrammetric process. In order to process these data and to obtain metrically

certified results, a modification of the algorithms of the standard photogrammetric pipeline was necessary. This purpose was achieved with the use of open-source Structure-from- Motion algorithms and the creation of a specific benchmark to compare the results.

Specifically, this thesis is divided into five Chapter. After the introduction in Chapter 1, Chapter 2 is dedicated to photogrammetry applied to historical images. A classification and a state of the art of historical archives material considering their possible use in metric documentation and thus suitable for photogrammetry is performed. Then, a photogrammetric workflow is proposed to process historical images and the maximum metric quality level reachable by the photogrammetric processing is investigated.

In Chapter 3 the state of the art of Deep Learning applied to Cultural Heritage is presented. In particular, an innovative match-moving method is proposed to improve ways to search for architectural heritage in video material and to reduce the effort of manually examining them by the operator in the archive in terms of efficiency and time.

Chapter 4 is concerned with the description of the case studies analysed and the discussion of the results of the implementation of the method. Two case studies in Paris were chosen: the UNESCO Heritage Tour Saint Jacques and the pavilions of Les Halles of the architect Victor Baltard. These case studies represent two different situations of heritage because the tower was transformed over time but still exists and the pavilions were destroyed in 1971. Thus, it is possible to compare the different results obtained from the implementation of the workflow to the two case studies. To validate the methodology, the workflow was tested also on other case studies and the results are reported: the historical photographs of paintings of Byzantine churches; a Japanese historic building before and after restoration and the temporary architecture of the International Exposition in Turin in 1921 and 1928.

Conclusions and future perspectives of the research will be provided in the final part of the thesis (Chapter 5).

Acknowledgment

First and foremost I would like to thank my supervisors, Prof. Fulvio Rinaudo and Prof. Rosa Tamborrino for their advice and continuous patience during my PhD study. Prof. Fulvio Rinaudo inspired my interest in the study of photogrammetry and guided the step of my research with constant and prompt feedback. He gave me the precious opportunity to travel during these years for attending external courses, presenting at conferences and collaborating with other scientists and researchers. Prof. Rosa Tamborrino offered me the possibility to attend the Summer School in Digital Humanities that she organized and this participation was instrumental in defining the application of my research. She also sustained me during the research period in Paris, giving support and suggestions. I am also extremely grateful for the stimulating experience to be a tutor at the Master course in Digital History hold by both the supervisors.

Besides my advisors, I would like to thank the rest of my thesis committee for their insightful comments and recommendations.

During my PhD path, I met a lot of esteemed researcher and professor that appreciated my work and give me the possibility to work with them. These experiences of course helped me to grow as a researcher and to improve my competencies. For this reason, following I would like to thank all these people that trust in me and my abilities.

I am extremely grateful to Francesco Salvadore and Stefano Tagliaventi, researchers at CINECA, to whom this thesis is dedicated. They taught me and supported me on the Artificial Intelligence and HPC part of this thesis. Since our first meeting, they believed in this project and accepted with enthusiasm to collaborate together. Their insightful feedback pushed me to sharpen my thinking

and brought my work to a higher level. Moreover, their immense knowledge and plentiful experience as scientists have encouraged me all the time to pursue my dreams and expectations in research and daily life. In the last years they were always there for me and showed on many occasions to be on my side, in the successful moment and in those more difficult. This was for me the most important thing and I really appreciated what they have done for me.

I would like to thank also the other members of the SCAI Department at CINECA in Rome that host me many times and gave me attention and care.

My sincere thanks go to Dr Ryo Higuchi and Prof. Satoshi Nasu of Tokyo Institute of Technology who provided me with an opportunity to join their team as visiting researcher, and who gave me access to the laboratory and research facilities. Without their precious support, it would not be possible to collect precious data for this thesis. I would like to acknowledge all the students of the lab that welcome me with enthusiasm and make me feel loved. I would also like to thank Prof. Hirofumi Sugawara of Kanazawa University for inviting me to visit him during my research period in Japan and proposing me a collaboration on his project. The results of it are part of this thesis. All these people have made my study and life in Japan a wonderful time.

I would like to extend my sincere thanks to Prof. Raphaële Héno for meeting me during my research period in Paris and for her assistance in the recovery material necessary for this thesis.

I would like to express thankfulness to the archive Lobster Films for sharing footage used in this research, to CNC, Forum des Images, Ina.fr, Gaumont Pathé Archives, Les Documents Cinematgraphique, and to ICONEM for kindly making available the model of the Tour Saint Jacques. I also express gratitude to the Cinema Museum and Bibliomediateca "Mario Gromo" of Turin for sharing the movie "Torino 1928" (Luis Bogino), and the historical archive of Politecnico di Torino's Library for the planimetry of "Exposition" held in Turin in 1928 and for the projects and photographs of "Mine and Ceramics" pavilion.

I acknowledge the CINECA award under the ISCRA initiative, for the availability of high-performance computing resources and support, and for giving me access to the I-Media-Cities during the experimentation phase.

I thank all my beloved PhD colleagues, and especially Marco and Marta, for the listening, attention and stimulating discussions, for living together with the difficulties and for all the fun we have had in the last four years.

I would like to offer my special thanks to all the friends that sustained me in different ways during the last years: my researcher-mate Valerio, Roberto, Daniele, my team-mate Antonia and Roberto; my advice-mate Alessio and Carmine (also for hosting me in Paris); my uni-mate Pierfrancesco and Manuela; my Salerno/Torino-mate Annarita, Sara, Martina; my crazy-mate Francesca, Vincenzo and Stefano.

Last but not the least, I would like to thank my family and myself for not giving up and for having the courage to go to the bottom of this path.

In memory of Stefano

To him and Francesco
For always been on my side

Contents

1. Introduction …………………………………………………………………… 1

1.1 Principle of transparency: correct use of metric information in the documentation process ………………………………………………………… 3
1.2 Motivation: Reconstructing lost or transformed Cultural Heritage with photogrammetry …………………………………………………………….. 5
1.3 Digital Humanities: Geomatics as support to historical studies ………..…... 8
1.4 The nexus of the research with Historical Archives ……………..…………. 9
1.5 Innovation of the research: Combining Artificial Intelligence and Photogrammetry ………………………………………………………………… 12

1.5. l Exploiting Artificial Intelligence for Cultural Heritage documentation 12
1.5.2 Open issues …………………………………………………………13
1.5.3 Aim of the thesis ………………………………………………….... 13
1.5.4 Proposed Workflow ……………………………………………….... 14
1.5.5 Structure of the thesis ……………………………………………… 15

References ……………………………………………………………………...19

2. Photogrammetry from historical images: Metric Quality Assessment 22

2.1 Processing historical images with photogrammetry 22

2.1. l Open issues in processing historical images 22
Issues related to archive storage ……………………………..................... 23
Issues related to the aim of the acquisition ……………………………..... 24
Issue related to technical features 24
2.1.2 Classification of historical images and state of the art 26
Single Images ………………………………………………………… 27
Photographs for Photogrammetry ………………………………….…... 28
Stereoscopic Views 28
Random Images 29
Videos 30
2.1.3 Structure-from-Motion pipeline with open source algorithms 30
2.1.4 Solutions proposed: COLMAP 32

2.2 Optimizing feature extraction in COLMAP 33

2.2.1 Proposed workflow 34
2.2.2 Feature point selection 35
2.2.3 Metric Quality Assessment: Point Cloud Comparison 37

2.3 A benchmark for historical film footage to assess the metric quality of the photogrammetric reconstruction 38

2.3.l Toward the need of a new video benchmark 38
2.3.2 Camera motions analysis and related works 40
Tilting 41
Trucking 41
Panning 42
2.3.3 Acquisition and processing of video dataset for the benchmark. …….. 42
Video Acquisition 43
Case study and dataset 44
Photogrammetric video processing in COLMAP 46
2.3.4 Metric quality assessment and evaluation results 50
Precision analysis 51
Accuracy analysis 59
References 66

3. Recovery material suitable for photogrammetry 75

3.1.1 Existing match-moving methods 77
3.1.2 Innovation of the proposed workflow 78

3.2 Architectural heritage detection using Neural Networks 79

3.2. l State of the Art: Artificial Intelligence for Cultural Heritage 80
3.2.2 Object Detection Using Neural Networks 81
Faster R-CNN Neural Network 81
SSD Neural Network …………………………………………………… 82
3.2.3 Evaluation Metrics of Neural Networks 84

3.3 Camera motions identification for photogrammetry 87

References 92

Appendix A 97

4. Applications on case studies and results 99

4.1 Case studies in Paris 103

4.1.1 Tour Saint Jacques 103
4.1.2 Les Halles 108
4.1.3 Recovering the material on the case studies 116
Historical video archives consulted 121

4.2 Testing the entire workflow: AI+ photogrammetry 123

4.2.1 Dataset 123
Dataset 1: Reference Case - Tour Saint Jacques 124
Dataset 2: Video 125
Dataset 3: Real Case-Tour Saint Jacques and Les Halles 127
4.2.2 Results and Discussion of the implementation of the Neural Networks on the case studies 129
Network Model Selection and Tuning 132
Assessment of the Training Dataset.. 138
Network Evaluation in a Realistic Scenario …….. 142
Frame-Based Metrics 142
Time Interval-Based Metrics. 143
Hardware Analysis: High-Performance Computing vs. General Purpose ……... 146
4.2.3 Camera motions identification 146
4.2.4 Photogrammetric processing and evaluation of metric quality assessment 151
Photogrammetric processing ……………………………………………… 152
Metric quality assessment - Benchmark comparison 154
Metric quality assessment - Cloud to mesh distance comparison 158
Metric quality assessment - Feature point comparison 160

4.3 Validation of the photogrammetric method …………………………….. 164

4.3.1 Photogrammetric reconstruction from historical photographs of restored Cultural Heritage ………………………………………………………................. 165
Churches in Goreme National Park and the Rock Sites of Cappadocia - Restored wall paintings …………………………………………………....... 165
Nakamachi Komise Street in Japan- Restored architecture ……………….. 166
Results of the application of the method to the two case studies ……..…. 167
4.3 .2 Photogrammetric reconstruction from historical film footage of ephemeral architecture ……………………………………………………………… 171
The Hungarian pavilion at Turin International Exposition of 1911 ……..... 172
The Mines and Ceramics pavilion at Turin International Exposition of 1928 ………………………………………………………………………….….... 172
Photogrammetric reconstruction of the pavilions and metric quality assessment ……………………………………………………………………………….. 174
References …………………………………………………………..................... 182

5. Conclusions …………………………………………………………….….… 184

5.1 Restating the aims of the dissertation ……………………………….…… 184
5.2 Originality and findings of the research …………………………………... 185
5.3 Applications and further research ……………………………………….... 186
5.4 Importance of this research for the present ……………………………. 188

List of Tables

Chapter 2

Table 1. Camera specifications …..……………………………………………………..……………… 43

Table 2. Correlation of the precision in terms of Standard Deviation (σ) related to the frame scale factor (m) and the value of the base ratio (B/Z) (Kraus and Waldhäusl, 1990) ………………………..………………………………………………. 51

Table 3. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding tilting camera motion case. ……………………………………..………………………………….. 52

Table 4. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding trucking camera motion case ……………..…………………………………………………….…….. 57

Table 5. Residuals between the measures of the same distances extracted from the point cloud of the laser scanner, chosen as reference, and the point cloud resulted from the video processing for the tilting camera motion ………………………………………………………………………..…….. 60

Table 6. Residuals between the measures of the same distances extracted from the point cloud of the laser scanner, chosen as reference, and the point cloud resulted from the video processing for the trucking camera motion …………………………………………………………..…………….. 63

Chapter 4

Table 1. Description of the training and validation dataset 1 on the reference case of the Tour Saint Jacques. For each training implementation (RUN A, RUN B and RUN C) different number and combination of images are used 125

Table 2. Description of the video dataset of the Tour Saint Jacques 125

Table 3. Description of the video dataset of Les Halles 126

Table 4. Combination and number of images from the dataset 3 used in each run of the training and validation phase on the real case of the Tour Saint Jacques 127

Table 5. Combination and number of images from the dataset 3 used in each run of the training and validation phase on the real case of Les Halles. 128

Table 6. Hardware comparison 146

Table 7. Accuracy values expressed in percentage according to different clusters nd thresholds 150

Table 8. Technical features of the films used during the photogrammetric processing 151

Table 9. Values of the benchmark, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding tilting camera motion case. The Tilting2 case, chosen as a reference in this processing concerning the Tour Saint Jacques for the comparison with the results of the frame extracted from the video "Études sur Paris", is highlighted in red 156

Table 10. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals of the results of the photogrammetric processing of the frame extracted from the video "Études sur Paris", compared with the Tilting2 case of the benchmark 156

Table 11. Values of the benchmark, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for

each photogrammetric processing, according to the corresponding trucking camera motion case. The Tilting3 case, chosen as a reference in this processing concerning Les Halles for the comparison with the results of the frame extracted from the video “La Destruction des Halles de Paris”, is highlighted in red.
.... 157

Table 12. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals of the results of the photogrammetric processing of the frame extracted from the video “La Destruction des Halles de Paris”, compared with the Trucking3 case of the benchmark
.... 158

Table 13. Residuals values resulted from the comparison of the distance between the 3D model by Iconem, the point cloud resulted from the photogrammetric processing of the video “Études sur Paris” and the drawing by the architect Gabriel Davioud
.... 160

Table 14. Distances values measured in project drawing of the architect Victor Baltard, the 3D model by the Commune de Nogent-sur-Marne, the description in Lemoine, 1980 and the point cloud resulted from the photogrammetric process of the video “La Destruction des Halles de Paris”
.... 163

Table 15. Distances values measured in project drawing of the architect Victor Baltard, the description in Lemoine, 1980 and the point cloud resulted from the photogrammetric survey of the pavilion’s cave in Yokohama (Japan)
.... 164

Table 16. Residuals values (ΔX , ΔY and ΔZ) of the feature points measured on the point cloud obtained from the processing of historical photographs (called Xold, Yold and Zold) and the point cloud of the recent photogrammetric survey (called Xnew, Ynew, Znew)
.... 171

Table 17. Technical feature of the film footage “Nei cantieri dell’esposizione”
.... 174

Table 18. Technical feature of the film footage "Torino 1928"
.... 175

Table 19. Values of the benchmark, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding trucking

camera motion case. The Trucking1 case, chosen as reference in this processing concerning the Hungarian pavilion for the comparison with the results of the frame extracted from the video “Nei cantieri dell’esposizione”, is highlighted in red
.... 178

Table 20. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals of the results of the photogrammetric processing of the frame extracted from the video “Nei cantieri dell’esposizione”, compared with the Trucking1 case of the benchmark
.... 178

Table 21. Values of the benchmark, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding trucking camera motion case. The Trucking4 case, chosen as reference in this processing concerning the “Mines and Ceramics” pavilion for the comparison with the results of the frame extracted from the video “Torino 1928”, is highlighted in red
.... 179

Table 22. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals of the results of the photogrammetric processing of the frame extracted from the video “Torino 1928”, compared with the Trucking4 case of the benchmark
.... 180

Table 23. Residuals values of distance measured on the project drawing and the point cloud of the “Mines and Ceramics” pavilion obtained from the photogrammetric process of video frame
.... 181

List of Figures

Chapter 1

Figure 1. (a) Albrecht Meydenbauer. (b) The first one of Meydenbauers photogrammetric camera that was invented in September 1858. (c) The French Cathedral in Berlin. One of Meydenbauers experimental photographs from 1882 (40 x 40 cm). 100 years later, between 1977 and 1982, the image was used for the reconstruction of the church which was severely damaged during World War II Albertz, 2001)

…..……………………………………………………………………………….. 7

Figure 2. The proposed workflow: the first step is the object detection using AI, the second step (used only in case of video) is the camera tracking, the third step is the 3D reconstruction following the SfM pipeline and the fourth step is the metric quality assessment

…..………………………………………………………………………………… 15

Figure 3. Tour Saint-Jacques la Boucherie (1508-22), Paris. (a) Henri Jean- Louis Le Secq, 1853, Musée Carnavalet. (b) Francesca Condorelli (author), 2019

….…………...………………………………………………………………… 17

Figure 4. Les Halles of the architect Victor Baltard (a) before the demolitions in 1971, Charles Marville, 1855, Musée Carnavalet / Roger-Viollet.. (b) the destruction of the Halles, Jean-Claude Gautrand, 1977, Galerie W

…..………………………………………………………………………………… 17

Figure 5. Historical archives photographs from G. de Jerphanion's work (1925-

42) and images of the actual state of the paintings in Karanlık Kilise (11th century)
….……………………………………………………………………………… 18

Figure 6. Pictures of the Former Matsuno-Yu building from local archive before the restoration in 2013
….……………………………………………………………………………… 18

Figure 7. Pictures of the Former Matsuno-Yu building from on-site survey in 2018 after the restoration
…..……………………………………………………………………………… 18

Figure 8. (a) The Hungarian pavilion at International Exposition in Turin in 1921 (Cornaglia, 2001). (b) the "Mines and Ceramics" pavilion at International Exposition in Turin in 1928, Library of Politecnico di Torino "Roberto Gabetti", 1928
…..……………………………………………………………………………… 18

Chapter 2

Figure 1. Historical photographs of the Valentino Castle, from the archive of GAM - Galleria Civica d'Arte Moderna e Contemporanea, in which some archive issues (bad conservation stage due to humidity) occur
…...……………………………………………………………………...… 23

Figure 2. Historical film footage of the Valentino Castle, from the I-Media- Cities platform and the video archive of Cinema Museum of Turin, in which some aim of acquisition issues (documentary) and technical features issues (low resolution quality and lack of film information) occur
…...……………………………………………………………………...… 24

Figure 3. Historical photographs of the Valentino Castle, from the archive of GAM - Galleria Civica d'Arte Moderna e Contemporanea – Fondo Gabino, in which some technical features issues (images of the same object are taken by different types of camera that differ in exposure and shooting) occur
…...………………………………………………………………………... 25

Figure 4. Classification into five main categories of historical images stored in historical archives, according to their aim of acquisition and their availability

…...………………………………………………………………..…………. 26

Figure 5. The pipeline described in this section 2.2 corresponds to the third step highlighted in red

…...……………………………………………………………………………... 34

Figure 6. Flowchart of the proposed workflow in which a step (in red) of "Feature point selection" was added to the standard SfM pipeline in COLMAP

..…………………..…………………………………………………………... 35

Figure 7. Workflow of the step of the Feature point selection algorithm

…...……………………………………………………………………………... 35

Figure 8. An example of image coordinates measured with WebPlotDigitizer tool and inserted in the software for the manual detection of Feature Points

…...……………………………………………………………………………... 36

Figure 9. An example of the database management in COLMAP in which key points are stored as row-major float32 binary blobs

…...……………………………………………………………………………... 36

Figure 10. Scheme of the three types of camera motions: Up/Down Motion-Tilting, Left/Right Motion-Trucking and Rolling Motion-Panning

…..……………………………………………………………………………... 40

Figure 11. The Valentino Castle, the seat of the Politecnico di Torino

………………………...……………………………………………………... 44

Figure 12. Frames extracted from the video sequences of the left wing of the courtyard of the Valentino Castle shot with tilting camera motion. The sequence 1 was shot at a distance of 43 m, the sequence 2 at 38 m, the sequence 3 at 33 m, the sequence 4 at 28 m, the sequence 5 at 18 m, the sequence 6 at 10 m, the sequence 7 at 5 m

……...…………………………………………………………………………... 45

Figure 13. Frames extracted from the video sequences of the façade of the Valentino Castle shot with trucking camera motion. The sequence 1 was shot at a distance of 85 m, the sequence 2 at 65 m, the sequence 3 at 45 m and the sequence 4 at 25 m

…...……………………………………………………………………………... 46

Figure 14. The result of the first step feature detection and extraction of the pipeline. The red points are the Feature Points automatically extracted from the software

………..…………………………………………………………………………... 47

Figure 15. The result of the second step feature matching and geometric verification of the pipeline. The green lines are a visualization of the rays that link two homologous points in two consecutive images

…..……………………………………………………………………………... 48

Figure 16. The result of the third step of reconstruction. The point cloud results are used for the metric quality assessment

…...……………………………………………………………………………... 48

Figure 17. Results of the three steps of the COLMAP SfM pipeline for the iterative reconstruction. Processing of the frame extracted from the video sequence of the left wing of the courtyard of the Valentino Castle shot with tilting camera motion. The sequence 1 was shot at a distance of 43 m, the sequence 2 at 38 m, the sequence 3 at 33 m, the sequence 4 at 28 m, the sequence 5 at 18 m, the sequence 6 at 10 m, the sequence 7 at 5 m

……………………………………...……………………………………... 49

Figure 18. Results of the three steps of the COLMAP SfM pipeline for the iterative reconstruction. Processing of the frame extracted from the video sequence of the façade of the Valentino Castle shot with trucking camera motion. The sequence 1 was shot at a distance of 85 m, the sequence 2 at 65 m, the sequence 3 at 45 m and the sequence 4 at 25 m

………………………...…………………………………………………... 50

Figure 19. Geometric sketch of two cameras perpendicular to the baseline and parallel each other (Kraus and Waldhäusl, 1990)

………………………………………...……………………………………... 51

Figure 20. An example of two consecutive frames for the tilting1 case in which the overlapping is over 90% and the precision dramatically decreases

…………………………………………………………...………………….. 53

Figure 21. Gaussian distribution of the Residuals values for the tilting1 case, video taken at a distance of 43 m

………………………………………….……………………………….. 53

Figure 22. Gaussian distribution of the Residuals values for the tilting2 case, video taken at a distance of 38 m

………………………………...…………………………………….….. 54

Figure 23. Gaussian distribution of the Residuals values for the tilting3 case, video taken at a distance of 33 m

…………………………………....…………………………………….. 54

Figure 24. Gaussian distribution of the Residuals values for the tilting4 case, video taken at a distance of 28 m

……………..…………………………………………………………….. 55

Figure 25. Gaussian distribution of the Residuals values for the tilting5 case, video taken at a distance of 18 m

……………………………………..…………………………………….. 55

Figure 26. Gaussian distribution of the Residuals values for the tilting6 case, video taken at a distance of 10 m

…………………………………....…………………………………….. 56

Figure 27. Gaussian distribution of the Residuals values for the tilting7 case, video taken at a distance of 5 m

………………………………...……………………………………….. 56

Figure 28. Gaussian distribution of the Residuals values for the trucking1 case, video taken at a distance of 85 m

………………………………...………………………………………….. 57

Figure 29. Gaussian distribution of the Residuals values for the trucking2 case, video taken at a distance of 65 m

…………………………..……………………………………………….. 58

Figure 30. Gaussian distribution of the Residuals values for the trucking3 case, video taken at a distance of 45 m

……………………...……………………………………………………….... 58

Figure 31. Gaussian distribution of the Residuals values for the trucking4 case, video taken at a distance of 25 m

…..…………………………………………………………………………... 59

Figure 32. Distances measured on the point cloud obtained from a laser scanner survey of the left wing of the Valentino Castle and used for the comparison with the point cloud of the same wing obtained from the video processing for the tilting camera motion case studies

…………………………...……………………………………………….... 59

Figure 33. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion tilting1

……………………………………………………………...…………….... 61

Figure 34. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion tilting2

…………………………………………………………………...……….... 61

Figure 35. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion tilting3

…...……………………………………………………………………….... 62

Figure 36. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion tilting4

…...……………………………………………………………………….... 62

Figure 37. Distances measured on the point cloud obtained from a laser scanner survey of the façade of the Valentino Castle and used for the comparison with the point cloud of the same façade obtained from the video processing for the trucking camera motion case studies

…………………………………………...……………………………….... 63

Figure 38. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion trucking1

……………………………………………………...…………………….... 64

Figure 39. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion trucking2 ………………………………………………………………………..……….. 64

Figure 40. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion trucking3 ………………………………………………………..………………….. 65

Figure 41. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion trucking4 ……………………………………………………………………………..….. 65

Chapter 3

Figure 1. The general workflow proposed in this dissertation. The following sections will deal with the application of this workflow on historical film footage, focusing on the two steps of "object detection using AI" and "camera tracking" ………………………..…………………..………………………………. 76

Figure 2. Workflow of the standard match-moving method that has been modified to improve it for a more efficient use for the heritage. A new workflow is proposed introducing the use of AI, the camera tracking algorithm, the use of open source SfM algorithms and the metric quality assessment of the results ………..……………………………………………………………….….. 79

Figure 3. Sketch of the Faster Region-based Convolutional Neural Network (Faster R-CNN) …………………………………………………………………………..... 82

Figure 4. Sketch of the Single Shot Detector (SSD) network ……………………………………………………………….…………. 83

Figure 5. Workflow of the second step of the proposed match-moving method: the tilting and trucking camera motions identification suitable for photogrammetry …………………..…………………………………………………………… 87

Figure 6. An example of the frame selected by the predict step of the algorithm.

…………………………………………………………………………………… 88

Figure 7. An example of the frame extracted by the image extraction step of the algorithm

…………………………………………………………………………………… 89

Figure 8. An example of the cluster of frame created by the frame clustering step of the algorithm

…………………………………………………………………………………… 89

Figure 9. An example of the centroid calculation by the algorithm

…………………………………………………………………………………… 90

Figure 10. Workflow of the camera motion estimation step of the algorithm

…………………………………………………………………………………… 91

Chapter 4

Figure 1. Paris in 1860, view of the Saint-Gervais Quarter, showing the location of the Tour Saint-Jacques at lower right, on the Seine's Right Bank and the Pavilions of Les Halles. From Philippe Benoist, Paris dans sa splendeur (Paris, 1861)

…………………………………………………………………………………… 100

Figure 2. View of Les Halles and the Tour Saint Jacques,1969, Archives de Paris

…………………………………………………………………………………… 100

Figure 3. Tour Saint-Jacques la Boucherie (1508-22), Paris. (a) Henri Jean-Louis Le Secq, 1853, Musée Carnavalet. (b) Francesca Condorelli (author), 2019

…………………………………………………………………………………… 101

Figure 4. Les Halles centrales by the architect Victor Baltard. (a) Pavilions 5 and 6, Charles Marville, 1855, Musée Carnavalet / Roger-Viollet. (b) Top view of Les Halles, Roger Henrard, 1952, Musée Carnavalet/ Roger-Viollet

…………………………………………………………………………………… 101

Figure 5. Maps and some buildings around Les Halles and the Tour Saint Jacques, Blondel La Rougery, 1959

…………………………………………………………………………………… 102

Figure 6. The Saint Jacques La Boucherie church. (a) Map of the city of Paris, Harold B. Lee, Scale: ca. 1:2.000, 1618, Library Maps Collection. (b) Plan of Turgot, 1734, Archives Nationales

…………………………...………………………………………………. 103

Figure 7. The Tour Saint Jacques. (a) The Saint Jacques La Boucherie church, Garnerey, 1784. (b) The tower before the isolation by restoration in the 1850s, "The Place du Chatelet and the Tour Saint-Jacques in 1848". (c) The Tour Saint Jacques on a decorative stone podium after relocation, Theodore Baldus, 1858. In O' Connell, 2001

…………...……………………………………………………………. 103

Figure 8. The Tour Saint Jacques, Henri Jean-Louis Le Secq, 1853, Musée Carnavalet.

……………………………….…………………………………………… 104

Figure 9. The Tour Saint Jacques during transformation works, view of Rue Saint-Martin and Maisons de la rue de la Vannerie, Martial, 1852, Bibliothèque nationale de France

…………………………...………………………………………………. 104

Figure 10. Plan Vasserot in which the Tour Saint Jacques is surrounded by buildings, Cadastre par îlots de Vasserot et Bellanger, 1810-1836, Archives Nationales de Paris

……………………...……………………………………………………. 105

Figure 11. Plan Jacoubet in which the Tour Saint Jacques is surrounded by buildings, 1836, Bibliothèque historique de la Ville de Paris

…..……………………………………………………………………. 105

Figure 12. Plan Haussmann Paris in which the Tour Saint Jacques was moved from its position and freed by sourroundings buildings, Plan parcellaire du prolongement de la rue de Rivoli Typ. Vinchon / Gráve par Avril Frères / Lith. Lemercier, 1851

………………..………………………………………………………. 106

Figure 13. The Tour Saint Jacques in the actual asset, Francesca Condorelli (author), 2019

…………………..…………………………………………………………………. 107

Figure 14. Les Halles. (a) Charles Marville, 1855, Musée Carnavalet / Roger-Viollet. (b) and (c) 1943, Musée d'Orsay

……………………………………..………………………………………… 108

Figure 15. Les Halles, Gilles Guérin, 1973, (http://www.gillesguerin.com/)

…………………………………………………………………………….….. 108

Figure 16. Les Halles, 1943, Musée d'Orsay and Musée Carnavalet / Roger-Viollet

……………...……………………………………………………………. 109

Figure 17. Details of Les Halles (in Thomine-Berrada, 2012)

…………………………………………………………...……...……. 110

Figure 18. Les Halles. (a) Detail of the roof. (b) The cave. (Medecine de France N° 226, 1971)

………………………………………………………...………………..... 111

Figure 19. The destruction of Les Halles (in Thomine-Berrada, 2012)

…………………………………………………………..……………. 111

Figure 20. The destruction of Les Halles, Jean-Claude Gautrand, 1977, Galerie W

…………………………………………………………………….....….... 112

Figure 21. Example of newspaper against the destruction of Les Halles, 1972, Musée d'Orsay

……………………………………………………………......…………. 112

Figure 22. Plan of Les Halles, 1867, Bibliothèque des Archives de Paris. The Pavilion No. 8 highlighted in red was moved in Nogent-sur-Marne and in Japan

……………………………………………………………......…………. 113

Figure 23. Pavilion n°8 in Nogent-sur-Marne, Francesca Condorelli (author), 2019

…………………………………………………………………..…………. 114

Figure 24. Pavilion No. 8 in Yokohama (Japan), Francesca Condorelli (author), 2019

……………………………………………………………………..………. 115

Figure 25. A diapositive of Les Halles at Musée d’Orsay, Francesca Condorelli (author), 2019

……………………………………………………………..…………. 116

Figure 26. Plan of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris

….……………………………………………………………………. 117

Figure 27. Façade of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris

…………………………………………………………..……………. 117

Figure 28. Section of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris

…………………………………….…………………………..…………. 118

Figure 29. Details of the cave of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris

…………….……………………………………………………………. 118

Figure 30. Details of the cave of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris

……………………………………...……………………………………. 119

Figure 31. dimensions of the 35 mm film format

………………..……………………………………………………………. 122

Figure 32. The workflow proposed in this dissertation

…………………..……………………………………………………. 123

Figure 33. A selection of the pictures from the dataset 1: (1) Tour Saint Jacques; (2) landscape; (3) negative matching; (4) tower parts

…………………………………..…………………………………………. 124

Figure 34. A selection of the pictures from the dataset 3 for the tower: (1) historical photographs; (2) historical images; (3) negative matching.

…..…………………………………………………………………………. 127

Figure 35. A selection of the pictures from the dataset 3 for Les Halles: (1) historical photographs; (2) historical images; (3) negative matching

……………………………………………………………………………. 128

Figure 36. A sketch of the metrics chosen to evaluate Neural Networks

……………………………………..………………………………. 129

Figure 37. Examples of frames True Positive of the Tour Saint Jacques: the monument was correctly detected by Neural Networks

……………………………………………………………………………… 130

Figure 38. Examples of frames True Positive of the Tour Saint Jacques: the monument was correctly detected by Neural Networks, even if it is a drawing

…………………………………………………………………………..… 130

Figure 39. Examples of frames True Positive of the Tour Saint Jacques: even the details and parts of the monument were correctly detected by Neural Networks

………………………………………………………..………………. 130

Figure 40. Examples of frames True Positive of Les Halles: the monument was correctly detected by Neural Networks

………………………………………..……………………………… 131

Figure 41. Examples of Les Halles: the monument was correctly detected by Neural Networks, even if it is a drawing

………………………..……………………………………………… 131

Figure 42. Examples of frames False Positive of the Tour Saint Jacques: the object was detected by Neural Networks as “Tour Saint Jacques” but it is not the tower

……………………………………………………………………………… 131

Figure 43. Examples of frames False Positive of Les Halles: the object was detected by Neural Networks as “Les Halles” but it is not the pavilion

………………………………..………………………………………… 131

Figure 44. Examples of frames False Negative of (a) the Tour Saint Jacques and (b) Les Halles: the objects were detected by Neural Networks as “No Tour Saint Jacques” and “No Halles” but they are the searched objects

…...…………………………………………………………………… 132

Figure 45. Examples of frames True Negative of (a) the Tour Saint Jacques and (b) Les Halles: the objects were correctly detected by Neural Networks as “No Tour Saint Jacques” and “No Halles” ………………………………………………………………………. 132

Figure 46. Sensitivity trend of RUN A with valid 1 and 4 …..…………………………………………………………………… 133

Figure 47. Specificity trend of RUN A with valid 2 and 3 …..…………………………………………………………………… 134

Figure 48. Sensitivity trend of RUN A and RUN B with valid 1 and 4, and threshold 0.9 …………………………..……………………………………………………… 135

Figure 49. Specificity trend of RUN A and RUN B with valid 2 and 3, and threshold 0.9 ……………………………………………..…………………………… 135

Figure 50. Sensitivity threshold analysis for RUN A and RUN B with valid 1 and 4 …..…………………………………………………………………… 136

Figure 51. Specificity threshold analysis for RUN A and RUN B with valid 2 and 3 ……………..………………………………………………………… 136

Figure 52. Sensitivity threshold analysis for RUN B and RUN C with valid 1 and 4 …………………………..……………………………………………… 137

Figure 53. Specificity threshold analysis for RUN B and RUN C with valid 2 and 3 ………………..……………………………………………………… 137

Figure 54. Real values of sensitivity analysis for T1 and T2 with valid5 and threshold equal to 0.9; (b) for H1 and H2 with valid1 and threshold equal to 0.9 …..…………………………………………………………………… 139

Figure 55. Real values of sensitivity analysis for H1 and H2 with valid1 and threshold equal to 0.9

…..…………………………………………………………………………… 139

Figure 56. Real values (red dots) and fitting curves (dotted line) of sensitivity analysis for T1 with valid5 and threshold equal to 0.9

…………………………………..…………………………………………… 140

Figure 57. Real values (red dots) and fitting curves (dotted line) of sensitivity analysis for T2 with valid5 and threshold equal to 0.9

……………………………….…………………………………………….. 140

Figure 58. Real values (red dots) and fitting curves (dotted line) of sensitivity analysis for H1 with valid1 and threshold equal to 0.9

………………………………..………………………………………….… 141

Figure 59. Real values (red dots) and fitting curves (dotted line) of sensitivity analysis for H2 with valid1 and threshold equal to 0.9

……………………………………………………………………………… 141

Figure 60. Sensitivity analysis evaluated on frames

…..…………………………………………………………………………… 143

Figure 61. Sensitivity analysis evaluated on frames

…...…………………………………………………………………...…. 143

Figure 62. Discovery Rate analysis

…..…………………………………………………………………………… 144

Figure 63. Time Save Rate analysis

…………………………………………..………………………………….. 145

Figure 64. Time Save Efficiency analysis

…………………………………………………………….…………….… 145

Figure 65.The workflow of the algorithm of camera motion identification

…..…………………………………………………………………………… 147

Figure 66. An example of the frames resulted from the predict step in which the Neural Network correctly detected the (a) Tour Saint Jacques, and (b) Les Halles

…………………………………………………………………………. 147

Figure 67. An example of the frames resulted from the frame extraction step in which the algorithm lists all the images selected by the predict step

……………………………………………..………………………………. 147

Figure 68. Example of the result of the frame cluster step according to the structural similarity threshold Ts. When the value of Ts is 0.1 there is a change of the scene

…………………..………………………………………………………… 148

Figure 69. Workflow of the camera motion estimation step of the algorithm

………………………………..…………………………………………… 148

Figure 70. Results of the camera motion estimation step that recognised the " camera tilting" movement

………………………………………………...………………………… 149

Figure 71. Results of the camera motion estimation step that recognised the "camera trucking" movement

…...……………………………………………………………………… 149

Figure 73. Results of the camera motion estimation step that recognised the "steady or oscillating camera" movement

……………………………………………………………..……………. 150

Figure 74. A selection of frames from the film footage "Études sur Paris" in which the Tour Saint Jacques appears shot with the tilting camera motion

……………………………………………………………...……………. 152

Figure 75. A selection of frames from the film footage "La Destruction des Halles de Paris" in which the Pavilion appears shot with the trucking camera motion..…………………………………………………………………. 152

Figure 76. The general workflow proposed in this dissertation in which the third step concerning the photogrammetric processing is highlighted in red and reported

Below

…..……………………………………………………………………… 153

Figure 77. A focus on the third step of the general workflow proposed in this dissertation in which the three steps of the photogrammetric pipeline are reported ……………………………………………………………………….……. 153

Figure 78. The frames identified from the film footage “Études sur Paris” with the object detection Neural Networks and the camera motion algorithm are selected and used for the photogrammetric process. The selection of the feature points, the results of the matching process and the final point clouds are obtained ………………………………………………….………………………… 153

Figure 79. The frames identified from the film footage “La Destruction des Halles de Paris” with the object detection Neural Networks and the camera motion algorithm are selected and used for the photogrammetric process. The selection of the feature points, the results of the matching process and the final point clouds are obtained …..…………………………………………………………………………… 154

Figure 80. The general workflow proposed in this dissertation in which the fourth step concerning the metric quality assessment is highlighted in red and reported below ……………………………..…………………………………………………… 154

Figure 81. A focus on the fourth step of the general workflow proposed in this dissertation in which the three steps of the metric quality assessment are reported. In particular the first stage of the analysis concerns the benchmark comparison and it is highlighted in blue …..…………………………………………………………………………… 155

Figure 82. Frames extracted from the video sequences 2 of the benchmark of the left wing of the courtyard of the Valentino Castle shot with tilting camera motion at a distance of 38 m, compared with the frame extracted from the video “Études sur Paris” …..…………………………………………………………………………… 155

Figure 83. Comparison of Normal Distribution of the Residual value between benchmark tilting2 case and case study of the Tour Saint Jacques …..…………………………………………………………………………… 156

Figure 84. Frames extracted from the video sequences 3 of the benchmark of the façade of the Valentino Castle shot with trucking camera motion at a distance of 45 m, compared with the frame extracted from the video “La Destruction des Halles de Paris”

……………………………………………….………………………. 157

Figure 85. Comparison of Normal Distribution of the Residual value between benchmark trucking3 case and case study of Les Halles

…..………………………………………………………………………. 158

Figure 86. A focus on the fourth step of the general workflow proposed in this dissertation in which the three steps of the metric quality assessment are reported. In particular the second stage of the analysis concerns the cloud to cloud (mesh) distance comparison and it is highlighted in blue

……………………………….…………………………………………. 159

Figure 87. Cloud to mesh distance comparison between the point cloud obtained from the photogrammetric processing of the video “Études sur Paris” and the 3D model obtained from the 3D survey by Iconem. (a) View of the entire Tour Saint Jacques. (b) A detail of the comparison

…..………………………………………………………………………. 159

Figure 88. A focus on the fourth step of the general workflow proposed in this dissertation in which the three steps of the metric quality assessment are reported. In particular the third stage of the analysis concerns the Feature Point comparison and it is highlighted in blue

………………………………..………………………………………… 160

Figure 89. Distances extracted from the drawing by the architect Gabriel Davioud of the Tour Saint Jacques

…………………………..……………………………………………… 161

Figure 90. Distances extracted from the drawing of Façade of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris

……………………………………………………………..…………………. 162

Figure 91. Point clouds of the Pavilion No. 8: (a) the point cloud has resulted from the photogrammetric processing of the film “La Destruction des Halles de Paris”. (b) the point cloud has resulted from a photogrammetric survey of the pavilion’s cave in Yokohama (Japan)

…..…………………………………………………………………………. 162

Figure 92. 3D model of the upper part of the Pavilion No. 8 made by the Commune de Nogent-sur-Marne, 2019

…..…………………………………………………………………………. 163

Figure 93. The general workflow proposed in this dissertation: for the two case studies analysed the validation was performed only in the last two steps: the SfM pipeline and the Metric quality assessment

…..…………………………………………………………………………. 164

Figure 94. Historical archives photographs from G. de Jerphanion’s work (1925-42) and images of the actual state of the paintings in Karanlık Kilise (11th century)

……………………………………………………………………..………. 165

Figure 95. Local archive photographs of the Former Matsuno-Yu building before the restoration in 2013

…..…………………………………………………………………………. 166

Figure 96. Images of the photographic on-site survey of the Former Matsuno-Yu building after the restoration in 2013, prof. Nasu, 2018

…..…………………………………………………………………………. 167

Figure 97. A focus on the third step of the general workflow proposed in this dissertation in which the three steps of the photogrammetric pipeline are reported

…………………………………………………………………….………. 167

Figure 98. Keypoints automatically extracted from the software with the standard “Feature detection and extraction” step: some important radiometric corners in the photograph are missing

…..…………………………………………………………………………. 168

Figure 99. Image coordinates of feature point measured with WebPlotDigitizer tool and added in COLMAP with the “Feature point selection” step

…………………………..………………………………………………. 168

Figure 100. Feature points manually chosen for the Karanlık church.

…...…………………………………………………………………..…….. 169

Figure 101. Feature points manually chosen for the Former Matsuno-Yu building

……………………………………..………………..……………….. 169

Figure 102. A focus on the fourth step of the general workflow proposed in this dissertation in which the three steps of the metric quality assessment are reported

……………………………...……………………………………………… 170

Figure 103. Cloud to cloud distance for the Karanlık church in which the point cloud obtained from the photogrammetric reconstruction from historical images is compared with the point cloud obtained from a photogrammetric recent survey of the church

……………………...……………………………………………………… 170

Figure 104. Cloud to cloud distance for the Former Matsuno-Yu building in which the point cloud obtained from the photogrammetric reconstruction from archival images is compared with the point cloud obtained from a photogrammetric recent survey of the building

…...………………………………………………………………………… 170

Figure 105. Pictures and postcards of the Hungarian pavilion in the Turin International Exposition of 1911

………………………………………………..………………………… 172

Figure 106. Planimetry of Turin Exposition 1928 – In red the “Mines and Ceramics” pavilion, Library of Politecnico di Torino "Roberto Gabetti", 1928

…………………………………………………………………...……… 173

Figure 107. Project drawing of “Mines and Ceramics” pavilion at Turin Exposition 1928. Plan, Section and architectural details, Pagano et al., 1930

……………………………………………..……………...……………… 173

Figure 108. The “Mines and Ceramics” pavilion at Turin Exposition 1928, Pagano et al., 1930

……………………………………………………………………………… 174

Figure 109. Frame from the film footage: “Nei cantieri dell’esposizione”, Museo del Cinema di Torino and I-media-cities, 1911

…...………………………………………………………………………… 175

Figure 110. A selection of the frames extracted from the video “Torino 1928” in which the pavilion of “Mines and Ceramics” appeared, Museo del Cinema di Torino and I-media-cities, 1928

…………………………...…………………………………...……………… 176

Figure 111. A focus on the third step of the general workflow proposed in this dissertation in which the three steps of the photogrammetric pipeline are reported

……………………………………………………………………………...…… 176

Figure 112. Feature detection and extraction, Feature matching and geometric verification, Point cloud resulted from structure and motion reconstruction of the

Hungarian pavilion from the film footage “Nei cantieri dell’esposizione”

…...………………………………………………………………………… 177

Figure 113. Feature detection and extraction, Feature matching and geometric verification, Point cloud resulted from structure and motion reconstruction of “Mines and Ceramics” pavilion from the film footage “Torino 1928

……………………………………………………………………...………… 177

Figure 114. Frames extracted from the video sequences 1 of the benchmark of the façade of the courtyard of the Valentino Castle shot with trucking1 camera motion at a distance of 38 m, compared with the frame extracted from the video “Nei cantieri dell’esposizione”

…...………………………………………………………………………… 178

Figure 115. Comparison of Normal Distribution of the Residual value between benchmark tilting2 case and case study of the Hungarian pavilion

………………………………………..………………………………….... 179

Figure 116. Frames extracted from the video sequences 4 of the benchmark of the façade of the courtyard of the Valentino Castle shot with trucking4 camera motion at a distance of 38 m, compared with the frame extracted from the video “Torino 1928”

………………………..……………………………………………….....… 179

Figure 117. Comparison of Normal Distribution of the Residual value between benchmark tilting2 case and case study of the “Mines and Ceramics” pavilion ………………………..…………………………………………………… 180

Figure 118. Measure extracted from the project drawing of the “Mines and Ceramics” pavilion. The distance AB was used to scale the model because present in the metric description of the pavilion ……………………………………………………………………………… 181

Chapter 1

Introduction

The thesis intends to give a specific contribution to the extraction of metric information from historical images and to experiment with its potentialities in the heritage field with the aim of valorising historical iconographical documentation to rise knowledge of historical buildings, with particular focus on the extreme case of lost or transformed monuments.

This thesis deals with historical images stored in historical archives in the heritage context. This documentation is part of the Intangible Cultural Heritage (ICH). According to the UNESCO definition:

> *"The "intangible cultural heritage" means the practices, representations, expressions, knowledge, skills – as well as the instruments, objects, artefacts and cultural spaces associated therewith – that communities, groups and, in some cases, individuals recognize as part of their cultural heritage. This intangible cultural heritage, transmitted from generation to generation, is constantly recreated by communities and groups in response to their environment, their interaction with nature and their history, and provides them with a sense of identity and continuity, thus promoting respect for cultural diversity and human creativity. For the purposes of this Convention, consideration will be given solely to such intangible cultural heritage as is compatible with existing international human rights instruments, as well as with the requirements of mutual respect among communities, groups and individuals, and of sustainable development.*" (*Convention for the Safeguarding of the Intangible Cultural Heritage, Art.1, 2003)*

This documentation has been produced for many other purposes but actually contains something very valuable for Cultural Heritage: data and information. Specifically, the thesis explores how to extract and use dimensional data from historical images for documenting monuments.

This thesis focuses on two kind of images documentation, historical photographs and film footage. For this reason, reference is made to the Digital Heritage (DH) according to the UNESCO definition:

"The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are "born digital", there is no other format but the digital object. Digital materials include texts, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained. Many of these resources have lasting value and significance, and therefore constitute a heritage that should be protected and preserved for current and future generations. This ever-growing heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression." (UNESCO Charter on the Preservation of the Digital Heritage, Art. 1, 2003).

Concerning the extraction of metric information from historical images, this thesis intends to give an upgrade of the previous studies on this topic using the latest developed technologies. This upgrade is in terms of metric precision provided with the use of classical photogrammetry combined with Artificial Intelligence solutions. The output is a method, suitable for different fields, but experimented with as an application in the heritage context.

The participation during the PhD path to courses and Summer Schools, such as the "Cities Cultural Heritage and Digital Humanities" organized by prof. Tamborrino in 2018, the research periods abroad in Paris and Tokyo in 2019, and the collaboration with CINECA to test the I-Media-Cities platform, have led to the selection of specific case studies used to experiment with the proposed methodology.

Historical images served as tests in a variety of situations, conditions and scales, depending on the different aspects that this study addressed: investigation of the maximum metric quality level reachable by the photogrammetric processing (Chapter 2), the test of the methodology on real cases of lost and transformed monuments (Chapter 4). The thesis also responds to issues related to the exploitation di digital collection proving a tool for the identification of specific monument in the image collection boosting the existing search engine with Deep Learning methods (Chapter 3). Finally, findings of the test and validation process of the proposed method are deeply discussed (Chapter 4 and 5).

1.1 Principle of transparency: correct use of metric information in the documentation process

The Cultural Heritage metric documentation plays an essential role in preserving memory and knowledge of the past and constitutes the set of information useful to plan any kind of interventions on Cultural Heritage assets.

The documentation is a common part of all the preservation, restoration, and management actions because provides all the information necessary to understand the object in question and leads to the adoption of best practices for the actions to be planned. The documentation information allows the virtual reconstruction of the investigated asset which forms today the starting point for each design of interventions and for complete knowledge about the present situation and, in case, of conditions of the same object in different historical epochs. All the documentation data represent a valuable source of knowledge that can be passed on to future generations and used for future actions (Stylianidis, 2019).

The knowledge of shapes and dimensions are one of the basic data of metric documentation useful also to locate all other non-metric information to help the comprehension of physical phenomena (e.g. structural diseases, etc.). The modern Geomatics techniques allow obtaining all these metric information with a certified accuracy and by extracting the best possible results by considering the quality of the used primary data.

The metric survey starts from the collection of a significant number of points with known coordinates in a unique reference system and the subsequent creation of a 3D model of the surveyed object. The 3D modelling requires a deep knowledge of the surveyed asset: as far as architectural assets (both single buildings and historical centres) the strong collaboration with specialists in History of Architecture is essential to correctly interpret and represent the original elements and the material interventions that occurred during the life of the investigated asset.

Rapid progress in the field of acquisition and processing of Cultural Heritage data has attracted many researchers and has led to a large number of studies in this direction. However, these rapid changes, while offer opportunities for collaboration between those who recover metric data (geomatics experts) and those who use them, at the same time they are presenting challenges to the professional partnership that may cause difficulty in the collaboration. Of particular concern is the issue related to the accuracy of the metric surveys. If this is not clearly specified by information providers, users of information may not recognise the importance and the limitation to their possible uses (Tucci and Lerma, 2018). Therefore, it is important to establish guidelines on how to take into account the different accuracy and precision requirements of 3D surveys according to the different end-user applications and purposes. Especially in the Cultural Heritage field, it is important to be aware that the metric surveying is not the end of a project but the starting point of many applications which strongly results in its metric quality. Geomatics experts play an important role in certifying the reliability of surveying and in supporting professionals who need metric information. Guaranteeing the quality of metric information extrapolated from 3D models ensures that they can be properly used for restoration projects and monitoring applications.

Recently, recovering the shape and dimensions of a Cultural Heritage asset has achieved characteristics of a mature and stable system with a well-defined system pipeline. Thanks to the development of advanced technologies in the field of geomatics and the use of low-cost devices, 3D surveying and modelling practices are becoming more and more fast and automatic. This phenomenon has led to numerous advantages, increasing the spread of these techniques to a wide platform of users (Tucci, 2016). However, the current trend to expand geomatics methods also for non-specialist includes the risk of not considering the required standards and the evaluation of quality and accuracy of metric data as essential information to properly manage the 3D model.

The concepts of ensuring that 3D metric survey methods are applied with scientific rigour are at the heart of several documents and papers which have emphasized the importance of accurate communication to users of the level of knowledge they represent.

Among these documents, certainly the London Charter represents an important step towards the definition of what is necessary for three-dimensional model generation in Cultural Heritage to be as rigorous and intellectually robust as any other research method. The purpose of the charter is to specify the goals and basic principles of the application of 3D metric survey methods concerning intellectual integrity, reliability, transparency, documentation, standards, sustainability, and access. The Charter aims to establish the principles which are sufficiently focused to have an impact, but abstract enough to remain current as to methods, and the technology continues to develop.

One of the most important issues for geomatics addressed in the document is concerned with the principle of “transparency”.

As declared in the section of the Charter dedicated to transparency requirements (Principle 4), this principle is based on providing enough information to allow a good understanding and evaluation of 3D visualisation methods and their results according to the contexts in which they are used and disseminated. The source, the type and the degree of uncertainty of the data and information collected must be specified. According to the objectives and the use of the method of 3D visualization, the type and quantity of transparency information will vary. The requirements for transparency of information may, therefore, vary from one project to another or at different stages of the same project. Moreover, to disseminate the documentation about the interpretation and decisions taken during the process it is really necessary to better understand the results. Each step of the process has to be documented in order to potentially reuse the data of the research, apply the results in different contexts, and guarantee the accessibility of information. Observing the issues highlighted in the Chart could be a solution to the transparency problem (Beacham et al., 2006).

The data that the London Charter requires to make the virtual model transparent can be provided only deriving them from a quality certification of all the steps of the 3D metric survey. However, a significant gap between theory and practice is still present and this represents a big challenge in the Cultural Heritage domain.

The core of the problem concerns the awareness that a model without this fundamental action of transparency does not reveal the effort in the process of creation. This "opacity" is in contrast with the need to show the documentation and reconstruction process, aspects that cannot be separated from the metric ones (Tucci, 2016).

1.2 Motivation: Reconstructing lost or transformed Cultural Heritage with photogrammetry

Among the various geomatic techniques, photogrammetry plays a fundamental role, since it allows the recovery of the metric data necessary for the geometric understanding of the object to be documented using only images. Recent developments in the acquisition and processing of photogrammetric data have reached a high level of automatism and the easy use of instruments and software has increasingly encouraged researchers and experts in the field of Cultural Heritage to use this technique in their works. Consequently, these advantages make it possible to extend the application of photogrammetric methods to low-skilled users and provide those who operate in the heritage fields with a tool for studying and intervening in Cultural Heritage.

However, automatism does not mean autonomy and therefore, while it is true that with any series of variously and randomly stereoscopic photos it is possible to obtain a 3D point cloud, it is not true that the quality is certifiable nor even less optimal. Simply the various automatic software does not care about to clearly state the level of precision reached during the process and even less the accuracy. Automatism has led to underestimating the metric survey. For example, it is increasingly common not to care about reducing overlapping to the minimum possible, to pay attention to the fact that inserting even just a divergent view the process could fail, to consider that photographs with strong scale differences threaten the metric quality and to care to understand the difference between precision (goodness of the measurements) and accuracy (correspondence of the survey to the real object).

For this reason, geomatics experts who create 3D models with high accuracy of analysis become indispensable support for those who use them for various applications. It is necessary to define the level of precision and accuracy achieved in order to enable the correct use of 3D models. Accuracy requirements are necessary to extract metric information and obtain high quality certified metric products that are essential for documentation. Recent developments in the field of photogrammetry for Cultural Heritage documentation have led to rapid progress in obtaining metric measurements. Of course in order to do that, it is necessary to acquire data with good primary quality (see CIPA 3x3 rules for example). Moreover, the need to perform a step-by-step check of results leads to control if the calibration of the images is acceptable or not, what is the variance-covariance matrix of the distortion calibration parameters in order to choose the most appropriate calibration model to use for relative image orientation, which are the suitable distances to use for the sizing of the point cloud that the software automatically generates.

Metric data without certified quality cannot provide the correct information and can lead to misuse. Researchers have shown more interest in this direction, and several studies have compared different acquisition tools and processing software in different situations and case studies. This continuous development also provides an opportunity to update well-defined approaches to extract metric information from images.

If this need for precision and accuracy assessment is important for new surveys, the same approach is necessary for the data extracted from historical information. In fact, one of the most fascinating challenges is the use not of new data but resources stored in historical archives.

Processing historical images, apart from some praiseworthy but very rare cases, the optimal conditions obtainable with images acquired ex-novo are rarely present and therefore it is all the more necessary to be able to verify the maximum result obtainable in terms of accuracy even more rigorously than in the case of images specially acquired for a photogrammetric survey.

However, archives are powerful platforms for saving invaluable treasures of enormous informative potential and play an essential role in the conservation of Cultural Heritage. In addition to written documents, old photographs and videos, which have been preserved over time, are in many cases unique witnesses to architectural and urban transformation. Monuments, historic buildings, and landscapes, that have been transformed or destroyed over time, appear in them and they become the only way to document changes of currently existing objects and parts that are no longer visible and to testify the state of buildings, parts of a city and urban environment at a specific time. It is obvious that terrestrial laser scanning (TLS) or photogrammetry on-site cannot be applied for buildings that do not exist anymore. In the case of assets that no longer exist, historical photographs and films footage are the only sources to recover their forms, dimensions and locations. This is an opportunity that could support historical studies and help in some way with restoration and conservation decisions.

In all international charters for the conservation of Cultural Heritage, photography is mentioned as one of the best ways of documenting cultural assets. These recommendations have always been interpreted as a need for photographic documentation, without taking into account the metric potential of photographic images and the benefit that these properties could provide for the effective documentation measures required before any kind of restoration.

Albrecht Meydenbauer, a young German architect, became a pioneer in the valorisation of Cultural Heritage through photogrammetry (Albertz, 2001).

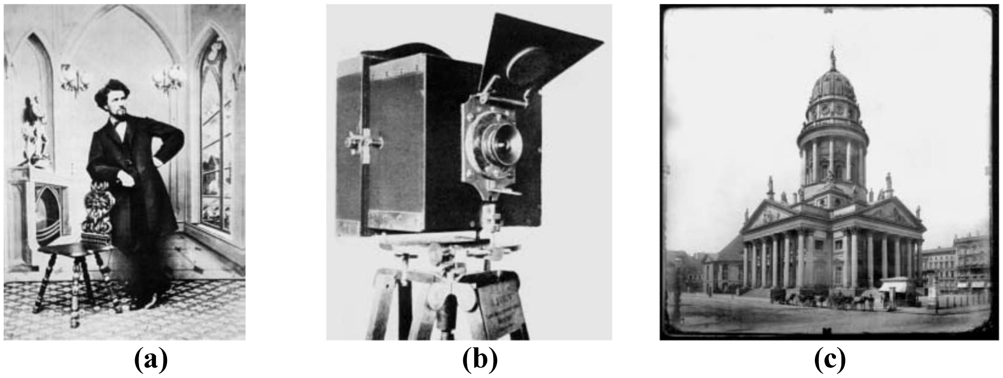


Figure 1. (a) Albrecht Meydenbauer. (b) The first one of Meydenbauers photogrammetric camera that was invented in September 1858. (c) The French Cathedral in Berlin. One of Meydenbauers experimental photographs from 1882 (40 x 40 cm). 100 years later, between 1977 and 1982, the image was used for the reconstruction of the church which was severely damaged during World War II (Albertz, 2001).

In 1858 he had the idea to use photographic images for the metric documentation of buildings. He was aware of the imminent danger to cultural assets and was convinced that the most important Cultural Heritage objects should be recorded in a Cultural Heritage Archive so that it could be reconstructed even if it was destroyed. Photogrammetric images were the most effective means to achieve this goal, and he had fought against many obstacles and criticisms to establish it as a method of scientific documentation. These images were widely used during the reconstruction of the city of Berlin at the end of the last century.

In recent years, thanks to the digitisation efforts of many archives, the interest in historical photographs and videos as valuable sources for the study of Cultural Heritage and the reconstruction of historical development has increased. The main problem with historical images is the availability of material of different types with low image quality, a total lack of camera parameter knowledge, the presence of deformations of the original dimensions, and damage due to improper storage.

However, recent developments in the field of image processing and Computer Vision have led to a renewed interest in processing data with a lack of essential properties for 3D metric information extraction. These improvements have increased the already demonstrated metric power of old images. In fact, image processing revisited from a photogrammetric point of view (search for homologous points) has given the possibility to automate the relative orientation and the generation of point clouds in a free and unscaled coordinate system. The projective geometry has made it possible to calibrate the distortions of photographic images concerning the central perspective in an automatic way (i.e.

without points of support). SfM is based on projective geometry while traditional photogrammetry is based on Euclidean geometry. However, the relative orientation made in projective geometry is less "robust" (statistically speaking) than the solution obtained in Euclidean geometry. In order to perform a complete metric analysis, the combination of the two approaches is a good solution. The projective geometry can be used for image calibration, automatic search for homologous points and relative orientation. Then traditional photogrammetry can improve the process and conclude it with the estimation of the mean and standard deviation of the internal and external orientation parameters and the coordinates of the homologous points.

1.3 Digital Humanities: Geomatics as support to historical studies

Digital 3D modelling and visualization technologies have been attracting a lot of interest not only in the geomatics field but also in the humanities research and education, especially but not exclusively on historical architecture (Münster, 2018). In recent years, in fact, have seen a growing trend towards finding new approaches to study humanities and to attract scholars in this field. The development of new informatics technologies has led the research to create powerful digital tools to support these studies. In this direction, certainly the Digital Humanities play a critical role in proposing systematic and technologically equipped methodologies in activities to support the traditional scholarly disciplines (Ganascia, 2015). Among these disciplines, architectural and urban history is a major area of interest within the field of Cultural Heritage studies. The need for preservation and documentation of historic cities has emerged as a promotion of the development of technologies to better analyse and understand historical sources. Examining and representing the past thanks to new communication tools allows the creation of a framework through the technology for people to experience, read, and follow an argument about a major historical problem (Seefeldt et al., 2009).

Digital humanities is not a discipline or a series of disciplines in itself. It is just a set of digital techniques that should help historical and literary research contributing to modify its research strategies. Digital history compared to history does not have different research objectives but uses different tools allowing more complete use of sources and the possibility to quickly manage large amounts of data to support research that is always the same.

One of the most important changes introduced by Digital Humanities is concerned with the combination of metric survey and keys of interpretation that allowed historians the match space and time in a more effective way (Tamborrino, 2014; Münster et al., 2019). The approach to historical research became more engaging thanks to the fruition of information on Cultural Heritage.

However, to apply digital techniques to Cultural Heritage studies, a close relationship between historians and technical experts is necessary. Encouraging collaboration in this multidisciplinary context is fundamental to create a common and interoperable language and thus, reaching high-quality results. As interdisciplinary collaborations are becoming more common, aligning the interests of geomatics and humanities scholars requires the formulation of a collaborative approach for research where the methodologies and intellectual innovations merge. Through data sharing, software sharing and knowledge sharing practices it is possible to improve this collaboration and to involve the participation of academic disciplines (Simeone et al., 2011). Boosting this cross-disciplinary cooperation is one of the missions of association such as ICOMOS/ISPRS CIPA, founded by the Cultural Heritage community with this intent.

To reach this objective certainly the construction of digital infrastructures such as digital libraries, archives, and repositories have produced new scenarios for collaborative research increasing exponentially the possibility of sharing data. The globally increasing digitization of data in private and public sectors improves the amount of information relevant to the researches in the field of Cultural Heritage and guarantees accessibility to contents and sources. For geomatics and urban historians, this creates an increasing interest especially for digital repositories dedicated to historic media such as photography and videos.

1.4 The nexus of the research with Historical Archives

The search for historical images suitable to recover metric information of a Cultural Heritage asset is a challenging problem to be addressed, both for geomatics and historians.

In this context, historical media archives have a central role. Images of the historic building are published in books, magazines, reviews and can be found in different status (negatives, positive copies, etc.) in public or private archives. Unfortunately, not all archives have a digital database to facilitate the search for interesting images. Digitised documents play an important role in the preservation of historical contents and their dissemination to the public. Without digital editions, the enormous amount of archives and old documents would not be easily accessible (Ioannides et al., 2017). Digitisation enables ubiquitous distribution, but also the enrichment of masterpieces with multimedia details and attractive content. Traditional simple Information Systems that support the management of cultural assets have given way to complex systems that display rich information from heterogeneous data sources (e.g. sensor networks, social networks, digital libraries, multimedia collections, web data service, etc.) employing sophisticated applications that enhance the user's experience (Amato et al., 2017). In this context, new digital technologies have facilitated the way to access Cultural Heritage information creating new tools to search, link and manage data. Moreover, facilitating links between information on Cultural Heritage, public perspectives and physical locations have enabled new ways of interacting with heritage and wider public participation.

One of the fundamental problems in dealing with Cultural Heritage data is to extract information from heterogeneous and unstructured sources. Consequently, there is a need to create innovative applications for exploring, analysing, mining and visualizing such data (Münster et al., 2016; Markhoff et al., 2017; Amato et al., 2017). Many libraries have understood the important role of the material they store and have recently started to share it with the public. However, open issues about historical material from archives (Münster et al., 2018) still exist (a detailed discussion is reported in Chapter 3). The first problem is the availability of the material, often made difficult by an enormous quantity of unorganized data on historic heritage. Archives can store a huge amount of data, so it often becomes impossible for scholars to evaluate and visually examine the archive manually. Consequently, a conventional Google search is often the main entrance for search queries (Beaudoin et al., 2011) and thus, organizing the archival data structure is a paramount challenge (Simeone et al., 2011). Especially, the filtering of data is still a persisting and challenging issue. Archival data are often tagged with metadata by varying operators after digitization, resulting in inconsistencies of metadata within a single archive and additionally, between multiple data repositories. While approaches for standardization of metadata exist (Lagoze et al., 2001; Elings et al., 2007), these are not strictly followed by every institution hosting an archive.

The second issue is concerning the rights of the data that if not licensed appropriately, the relevant data are not available for research. Sometimes even the status of the ownership is unknown and e.g. images may not be reproduced, changed or published. A Creative Commons licensing of the data is often available only for a very small part of the complete archive. Additionally, the quality of the digital copy is not consistent and improving the resolution requires funding. Since the digitization in archives is not yet completed the process of digitization remains the most important step for the data providers and institutes (Pandey et al., 2014).

The third issue is related to the involvement of architectural historians and skilled researchers in the design of digital media platforms. Generally libraries, archives and museums are responsible for images repositories and search tools and do not usually meet the requirements of architectural history scholars. In architectural historical research, the visual character of the sources instead of texts and words is essential. This could be a problem if an ordinary building or specific architectural features have to be searched by scholars. Moreover, since the existing applications for searching platforms are devised by computer scientist, the degree of search expertise of the operator in archives dramatically influences the success of the use of such tools (Friedrichs et al., 2018).

Recently, the increasing availability of photographs and videos in digital format has led to new research for the exploitation of the archives. It has provided the experiment of a web platform for access to this digital content. The huge amount of data stored in historical archives makes it desirable that their annotation and analysis have been automated and require minimal user intervention. For this reason, previous research has conducted in order to facilitate the access and the use of digital information.

One recent example is I-Media-Cities (https://imediacities.eu, May 2020). It is an innovative research project carried out by 9 European cultural institutions (film and audio-visual archives from 8 different countries) to share, access and use their digital content, making it a lever for new approaches in multidisciplinary research, for innovation in the economy and the general accessibility of the European Cultural Heritage. As an experimental innovation action, the project focuses on digital content related to cities. I-Media-Cities aims to be a cross-border and multilingual platform for the study of the history and urban development of large EU cities by providing large collections of media that are generally not easy to access, and for the study of the history of media through the way they have represented urban spaces. The project provides a platform for access to digital content (interoperable and multilingual) and makes it available to a growing community of researchers and creators across Europe. Photographs and videos can be searched as they are automatically annotated using a Machine Learning approach that makes them searchable through dynamic maps and semantic searches (Caraceni et al., 2017).

The previous example of similar platform exist and are reported here below.

PHAROS, The International Consortium of Photographic Archives, is an international consortium of fourteen European and North American art historical photo archives committed to creating an open and freely accessible digital research platform allowing for comprehensive consolidated access to photo archive images and their associated scholarly documentation. The PHAROS collections collectively contain an estimated 25 million images documenting works of art and architecture and the history of photography itself. The Getty Research Institute is one of the partners, and in collaboration with four international advisory institutions, led a project to create an online search platform that would unify and provide global access to digitized art history books and journals, including fundamental texts, rare books, exhibition catalogues, auction sales catalogues, and related literature. Launched in spring 2012, the Getty Research Portal is a trusted destination for researchers worldwide and a tool to assist librarians in planning future digitization projects (Salomon, 2014). CLIOH (Cultural Digital Library Indexing our Heritage) is a video indexing and retrieval system for an archaeological database using self-organizing neural networks (Huang et al., 2002). The CLARIAH Media Suite (2015) is one of the applications of the Dutch infrastructure for Digital Humanities and Social Sciences developed in the CLARIAH project that aims at the realisation of a common infrastructure for the humanities and social sciences. It facilitates access to key Dutch media collections with advanced multimedia search and analysis tools and user-friendly applications for the processing of these data. Another example is the online database Cinema Context (www.cinemacontext.nl, May 2020), a relational database and research instrument for studying the history of film culture in the Netherlands, created as structured datasets relating to the contexts of film production, distribution and consumption (Noordegraaf et al., 2018). The

ArchiMediaL project (http://archimedial.eu/, May 2020) in close cooperation between architectural historians and computer scientists experiments with the automatic recognition of architectural and urban forms in diverse visual media that are available digitally or on the web. The aim is to solve the metadata problem providing a way to search through these huge collections with descriptive keywords by using Artificial Intelligence solutions which can identify and correctly label descriptive data in pictures and paintings without an expert in art and architecture knowledge (Brouwer, 2018).

In all these examples has emerged that Machine Learning has a pivotal role in the advances of the process of the search for within archives material and in the following section, the potentialities of this method are highlighted and deepened.

1.5 Innovation of the research: Combining Artificial Intelligence and Photogrammetry

1.5.1 Exploiting Artificial Intelligence for Cultural Heritage documentation

Thanks to the effort of cultural institutions such as museums, galleries and heritage management organizations in investing a great deal of resources to digitize and preserve their collections using state-of-the-art acquisition technologies, this process have often been considered a success. Multiple initiatives such as high-quality replicas of cultural objects, virtual museum tours, digital valorisation, etc. have developed a new cultural and systemic awareness of the importance of data on Cultural Heritage.

With the recent hype in the field of Artificial Intelligence (AI), new techniques have been developed to manage them with Machine Learning (ML) and Deep Learning (DL) provide tools to decision-makers. In the past, cultural data enrichment was only possible using manual annotations that did not fully exploit the hidden information that could be extracted with AI technologies. Today, new challenges have arisen for researchers to make the digital preservation of assets more efficient with Artificial Intelligence techniques for content classification and generation.

One of the most successful aspects of the spread of AI is its application in several disciplines. Artificial Intelligence, in fact, involves, for example, computer science, engineering, art, medicine, linguistics etc. The blending of disciplinary fields is also the starting point for a cultural change that no longer differentiates between humanities, science and art disciplines. (Andrianaivo et al., 2019).

Since the lack of involvement of the researcher in the humanities in the design of infrastructures on historical heritage material, the application of AI methods in urban and architecture projects can improve the participation of the final users of such tools.

In fact, for what concerns the documentation process and in particular the collection of data and information about heritage, can really be improved if

Artificial Intelligence is combined with techniques widely used in the heritage field such as photogrammetry.

For this reason, recent research in this field has seen a rapid development of technologies to support the management and analysis of historical heritage data. Through Artificial Intelligence, tasks such as processing these large amounts of data and reducing human effort can be automated and thus made more efficient.

The creation of new tools for the end-user of these data is an interesting research topic, especially in the field of Cultural Heritage. Indeed, the volume, the size and the variety of historical data lead to certain critical factors. The most important of these is the manpower required to organise and search for the documents.

To solve this problem, the application of Artificial Intelligence offers the possibility to enhance historical archives and the retrieval of information on Cultural Heritage.

1.5.2 Open issues

Summarizing what explained in the previous sections, some issues to address have been raised.

First of all, the principle of transparency has emerged as the requirement to certify metrically the results of the documentation process following a geomatics approach. This general principle was adopted as an international standard now in use. However, the determination of metric accuracy is technically challenging when dealing with historical images from archives. Another important issue, in fact, concerns the limitations in processing historical photographs and film footage, since sometimes they have characteristics not suitable for the standard photogrammetric workflows because acquired not for this purpose. Furthermore, it was also highlighted that the difficulty of finding the material which often requires physical access to the archives, since in many cases it allows on-site consultation but not data sharing, is somehow resolved by the development of international projects (such as I-Media-Cities and others already mentioned) aimed at limiting the barriers to access to data in video archives. However, a fundamental problem that remains unsolved is the need to identify the object of interest within the amount of material that potentially contains it. The indexing of metadata for historical archival material is often incomplete or inaccurate, and the corresponding search engines are therefore not very efficient. The human effort to find the data of interest represents a significant percentage of the geomatics specialist and final user work.

1.5.3 Aim of the thesis

The specific objective of this thesis is to offer an analysis and assessment through the metric potentialities of different images available in historical archives, by considering the essential role of photogrammetry to extract metric information and to obtain a 3D model. The aim is to explore how metric information about the scale of buildings and groups of buildings, which no longer exist or transformed over time, could be extracted from early 19th century photographs and videos of different quality, for 3D virtual reconstruction analysing the material stored in historical archives to support researchers and experts in historical research of Cultural Heritage.

In order to process these data and to obtain metrically certified results, a modification of the algorithms of the standard photogrammetric pipeline was necessary. This purpose was achieved with the use of open-source Structure-from-Motion algorithms and the creation of a specific benchmark to compare the results.

Besides the processing of historical photograph, photogrammetry is combined with Artificial Intelligence to improve ways to search for architectural heritage in video material and to reduce the effort of manually examining them by the operator in the archive in terms of efficiency and time.

1.5.4 Proposed Workflow

In the workflow proposed in this work, a combination of Deep Learning techniques with photogrammetry is presented. DL is used for the retrieval of primary data used as input material in the standard Structure-from-Motion (SfM) pipeline used to reconstruct lost Cultural Heritage.

In particular, the first step of the workflow was to use object detection Neural Networks trained to automatically recognise the monument in film footage and image collections.

In the second stage of the workflow, specifically for the video, the frames suitable to be processed with photogrammetry are selected from all the frames detected by the Neural Network. The selection is performed according to the camera motions within the scene of the video. Only the shots taken from multiple points of view of the same scene are suitable for the photogrammetric process.

The third step concerned the photogrammetric reconstruction of the heritage with open-source algorithms in COLMAP developed by ETH of Zurich, (COLMAP, Johannes L. Schoenberger, 2019). During the process specific feature points are manually selected in order to guarantee their presence in the final point cloud.

Finally, during the fourth of the metric quality assessment of the model, the results of the 3D reconstruction of the heritage were compared with a benchmark specifically created to evaluate the metric quality of the model according to the type of camera motion used.

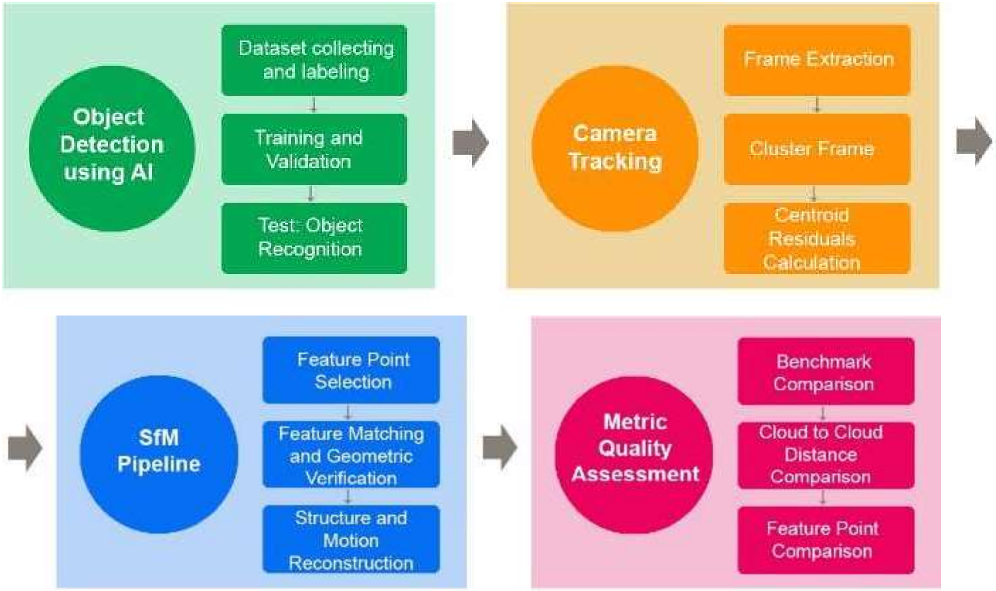

Object Detection using AI
Dataset collecting and labeling
Training and Validation
Test: Object Recognition
Camera Tracking
Frame Extraction
Cluster Frame
Centroid Residuals Calculation
SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction
Metric Quality Assessment
Benchmark Comparison
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 2. The proposed workflow: the first step is the object detection using AI, the second step (used only in case of video) is the camera tracking, the third step is the 3D reconstruction following the SfM pipeline and the fourth step is the metric quality assessment.

1.5.5 Structure of the thesis

This thesis is divided into four parts. A brief description of each part is reported here below.

The first part is dedicated to photogrammetry applied to historical images. A classification of historical archives material considering their possible use in metric documentation and thus suitable for photogrammetry is performed. For each category recognized, technical problems and criticality on processing these data are deeply treated since it represents one of the most studied topics, especially in the field of documentation of Cultural Heritage. In particular, a state of the art concerning the previously proposed methods to extract metric information from historical material is presented. After a description of the advantages of the use of open-source software for photogrammetry, the choice of the one suitable to process these kinds of data in this dissertation is explained.

Then, the two innovative methodologies to assess the metric quality of the photogrammetric results are proposed. The first one, suitable for both type of historical images, consists in improving the performance of open-source SfM algorithms in order to guarantee the presence of strategic feature points in the resulting point cloud, even if sparse. To reach this objective, a photogrammetric workflow is proposed to process historical images. The first part of the workflow presents a method that allows the manual selection of feature points during the photogrammetric process. The second part evaluates the metric quality of the reconstruction based on a comparison with a point cloud that has a different density from the sparse point cloud.

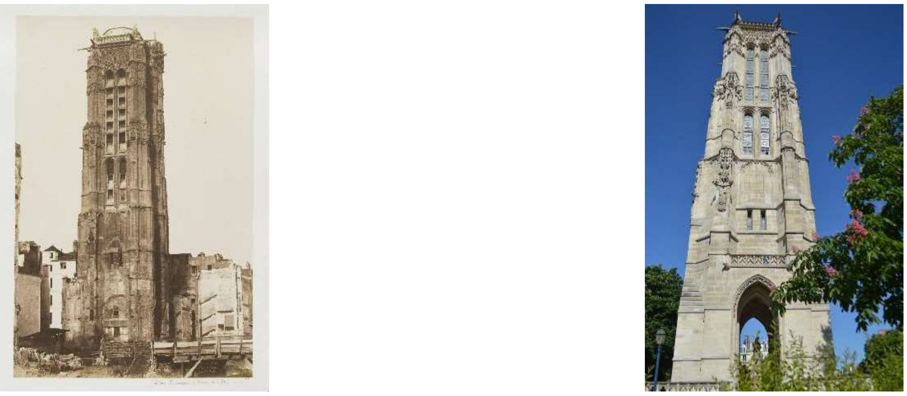
The second one, more specific for video, concerns the creation of a benchmark to evaluate the maximum metric quality reachable from this kind of processing. A new video dataset was collected with the aim of reproducing the camera motions in which the old video was shot. Three different camera motions were considered: Up/Down Motion-Tilting, Left/Right Motion-Trucking and Rolling Motion-Panning. The methodology was experimented on Valentino Castle in Turin, a monument inscribed in the UNESCO World Heritage List. Data were processed with the implementation of open-source Structure-from-Motion algorithms and the results were analysed for the evaluation of metric quality. Results show the different maximum precision assessments according to the different typologies of camera motion.

In the second part, the state of the art of DL applied to Cultural Heritage is presented. In particular an innovative match-moving method is proposed that aims to exploit Artificial Intelligence and SfM algorithms to identify the frames extracted from film footage in which the lost monument appears and that is suitable to be processed with photogrammetry for its 3D reconstruction.

This part is divided into two sections. In the first one, the open issues in collecting historical material are identified. An algorithm implemented to automatically detect monument in video sequences is also described here. The choice and the tuning of Neural Networks, the methodology used for testing them are described, with particular focus on the new metrics introduced for the evaluation of the algorithm in a real case. Following the description of the collection and preparation of the datasets are reported. The second section deals with the identification of video frames suitable to be processed with photogrammetry according to different types of camera motions.

The third part is concerned with the description of the case studies analysed and the discussion of the results of the implementation of the workflow.

Two case studies in Paris were chosen: the UNESCO Heritage Tour Saint Jacques (Figure 3) and the pavilions of Les Halles of the architect Victor Baltard (Figure 4). These case studies represent two different situations of heritage because the tower was transformed over time but still exists (the study is focused on the tower after transformations) and the pavilions were destroyed in 1971. Thus, it is possible to compare the different results obtained from the implementation of the workflow to the two case studies.



(a) (b)

Figure 3. Tour Saint-Jacques la Boucherie (1508-22), Paris. (a) Henri Jean-Louis Le Secq, 1853, Musée Carnavalet. (b) Francesca Condorelli (author), 2019.

(a) (b)

Figure 4. Les Halles of the architect Victor Baltard (a) before the demolitions in 1971, Charles Marville, 1855, Musée Carnavalet / Roger-Viollet.. (b) the destruction of the Halles, Jean-Claude Gautrand, 1977, Galerie W.

The methodology and the quality of the results were analysed, with particular focus on each part of the workflow previously described.

To validate the methodology, the workflow was tested also on other case studies and the results are reported: the historical photographs of paintings of Byzantine churches (Figure 5); a Japanese historic building before and after restoration (Figure 6 and 7) and the temporary architecture of the International Exposition in Turin in 1921 and 1928 (Figure 8).

Conclusions and future perspectives of the research will be provided in the final part of the thesis.

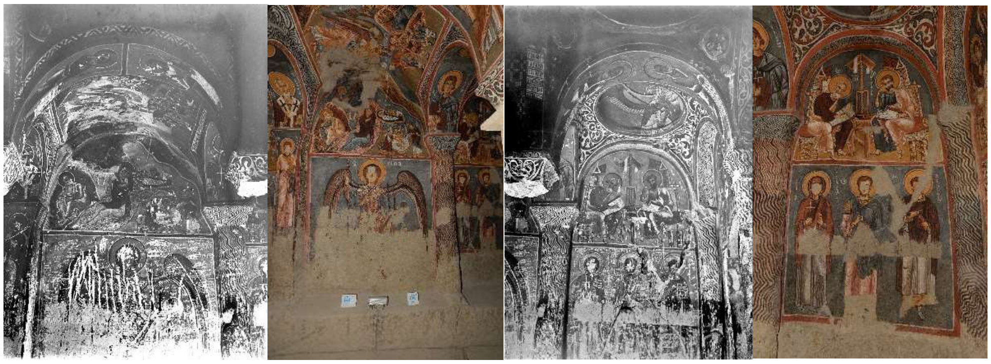


Figure 5. Historical archives photographs from G. de Jerphanion’s work (1925-42) and images of the actual state of the paintings in Karanlık Kilise (11th century).

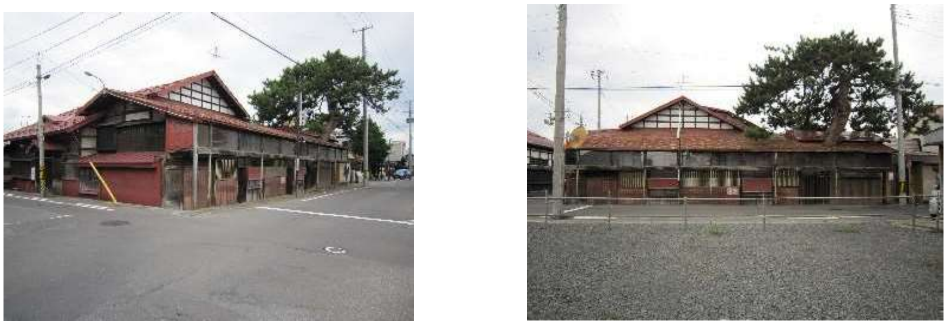


Figure 6. Pictures of the Former Matsuno-Yu building from local archive before the restoration in 2013.

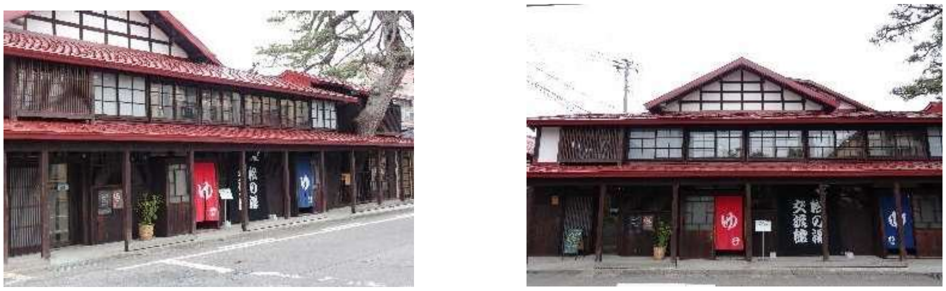
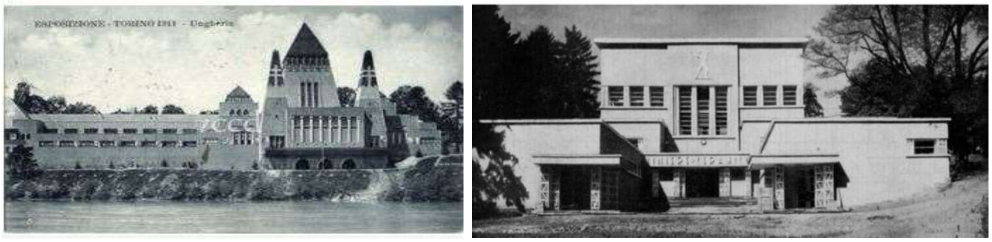


Figure 7. Pictures of the Former Matsuno-Yu building from on-site survey in 2018 after the restoration.



(a)

(b)

Figure 8. (a) The Hungarian pavilion at International Exposition in Turin in 1921 (Cornaglia, 2001). (b) the “Mines and Ceramics” pavilion at International Exposition in Turin in 1928, Library of Politecnico di Torino "Roberto Gabetti", 1928.

References

Albertz, J., 2001. Albrecht Meydenbauer – Pioneer of photogrammetric documentation of the Cultural Heritage. In: Proceedings 18th International Symposium CIPA 2001, September 18 - 21, 2001, Potsdam, Germany.

Amato, F., Moscato, V., Picariello, A., Colace, F., Santo, M.D., Schreiber, F.A., Tanca, L., 2017. Big data meets digital Cultural Heritage: Design and implementation of scrabs, a smart contextaware browsing assistant for cultural environments. Journal on Computing and Cultural Heritage (JOCCH) Vol. 10 No. 1, 6 (2017).

Andrianaivo, L. N., D'Autilia, R., and Palma, V., 2019. Architecture recognition by means of convolutional neural networks, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W15, 77–84, https://doi.org/10.5194/isprs-archives-XLII-2-W15-77-2019, 2019.

Beacham, R., Denard, H., Niccolucci, F., 2006. The E-volution of ICTechnology in Cultural Heritage, Papers from the Joint Event CIPA/VAST/EG/EuroMed Event, 2006 An Introduction to the London Charter.

Beaudoin, J.E., Brady, J.E., 2011. Finding visual information: a study of image resources used by archaeologists, architects, art historians, and artists. Art Documentation: Journal of the Art Libraries Society of North America Vol. 30 No. 2, 24-36 (2011).

Brouwer, A. 2018. Crowdsourcing architectural knowledge: Experts versus non-experts.

Caraceni, S., Carpenè, M., D'Antonio, M., Fiameni, G., Guidazzoli, A., Imboden, S., Liguori, M. C., Montanari, M., Trotta, G., Scipione, G., Hanegreefs, D, 2017. I-media-cities, a searchable platform on moving images with automatic and manual annotations. 23rd International Conference on Virtual System & Multimedia (VSMM), Dublin, Ireland. doi.org/10.1109/VSMM.2017.8346274.

Elings, M.W., Waibel, G., 2007. Metadata for all: Descriptive standards and metadata sharing across libraries, archives and museums. First Monday Vol. 12 No. 3, (2007).

Friedrichs, K., Münster, S., Kröber, C., and Bruschke, J., 2017. Creating Suitable Tools for Art and Architectural Research with Historic Media Repositories. In: Digital Research and Education in Architectural Heritage, pp. 117-138.

Ganascia, J.-G., 2015. The logic of the big data turn in digital literary studies. frontiers in Digital Humanities Vol. 2, 7 (2015).

Héno, R, Egels, Y., Heipke, C., Grussenmeyer, P., 2012. An overview of close-range photogrammetry in France. In: Revue Francaise de Photogrammetrie et de Teledetection n° 200, pp. 80-87.

Huang J., Umamaheswaran D., Palakal M., 2002. Video Indexing and Retrieval for Archeological Digital Library, CLIOH. In: Lew M.S., Sebe N., Eakins J.P. (eds) Image and Video Retrieval. CIVR 2002. Lecture Notes in Computer Science, vol 2383. Springer, Berlin, Heidelberg.

Ioannides, M., Davies, R., Chatzigrigoriou, P., Papageorgiou, E., Leventis, G., Nikolakopoulou, V., Athanasiou, V., 2017. 3D Digital Libraries and Their Contribution in the Documentation of the Past. In: Ioannides M., Magnenat-Thalmann N., Papagiannakis G. (eds) Mixed Reality and Gamification for Cultural Heritage. Springer, Cham. https://doi.org/10.1007/978-3-319-49607-8_6.

Lagoze, C., Van de Sompel, H., 2001. The Open Archives Initiative: Building a low-barrier interoperability framework. In: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries, pp. 54-62. ACM (2001).

Markhoff, B., Nguyen, T.B., Niang, C., 2017. When it comes to querying semantic Cultural Heritage data. In: European Conference on Advances in Databases and Information Systems, pp. 384-394. Springer (2017).

Münster S., 2018. Drawing the "Big Picture" Concerning Digital 3D Technologies for Humanities Research and Education. In: Chowdhury G., McLeod J., Gillet V., Willett P. (eds) Transforming Digital Worlds. iConference 2018. Lecture Notes in Computer Science, vol 10766. Springer, Cham.

Münster, S., Kamposiori, C., Friedrichs, K., Kröber, C., 2018. Image libraries and their scholarly use in the field of art and architectural history. International Journal on Digital Libraries Vol. 19 No. 4, 367-383 (2018).

Münster, S., Pfarr-Harfst, M., Kuroczyński, P., Ioannides, M., 2016. 3D Research Challenges in Cultural Heritage II: How to Manage Data and

Knowledge Related to Interpretative Digital 3D Reconstructions of Cultural Heritage. Springer (2016).

Münster, S., Apollonio, F. I., Bell, P., Kuroczynski, P., Di Lenardo, I., Rinaudo, F., and Tamborrino, R., 2019. Digital cultural heritage meets digital humanities, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W15, 813–820, https://doi.org/10.5194/isprs-archives-XLII-2-W15-813-2019, 2019.

Noordegraaf, J., Lotze (Kathleen Lotze), K., & Boter, J., 2018. Writing Cinema Histories with Digital Databases: The Case of Cinema Context. Tijdschrift voor Mediageschiedenis, 21(2), 106-126.

Pandey, P., Misra, R., 2014. Digitization of library materials in academic libraries: Issues and challenges. Journal of Industrial and Intelligent Information, vol. 2, (2014).

Salomon, K., 2014. Facilitating Art-Historical Research in the Digital Age: The Getty Research Portal. Getty Research Journal, Volume 6, Number 2014, https://doi.org/10.1086/675796.

Seefeldt, D., Thomas III, W.G., 2009. What is digital history? A look at some exemplar projects.

Simeone, M., Guiliano, J., Kooper, R., Bajcsy, P., 2011. Digging into data using new collaborative infrastructures supporting humanities-based computer science research. First Monday Vol. 16. No. 5, (2011).

Stylianidis, E., 2019. CIPA - Heritage Documentation: 50 Years: Looking Backwards, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W14, 1–130, https://doi.org/10.5194/isprs-archives-XLII-2-W14-1-2019, 2019.

Tamborrino, R., 2014. Digital urban history: telling the history of the city in the age of ICT revolution. Università degli studi Roma Tre, CROMA, Rome.

Tucci, G., Lerma, J.S., 2018. Special Issue GEORES2017. Geomatics and Restoration: Conservation of Cultural Heritage in the Digital Era. Applied Geomatics, Volume 10, Issue 4, 277–278.

Tucci, G., 2016. Geomatica e patrimonio digitale. Dai dati ai modelli: quale futuro?, Bollettino SIFET n.3 – ANNO2016: Sezione Scienza.

Chapter 2

Photogrammetry from historical images: Metric Quality Assessment

2.1 Processing historical images with photogrammetry

2.1.1 Open issues in processing historical images

The determination of the metric quality of the 3D models is technically challenging when derived from historical archives data. Assessing this quality is very important because historical documents such as images, maps, plans are the principal sources in the field of research of Digital Humanities. Among these materials, certainly the historical images represent a great potential because they can be processed with photogrammetry to extract metric knowledge to document for restoration and virtual reconstruction of destroyed or transformed objects. Orienting and registering historical pictures can also give a still unknown view of buildings and details that were not only available with two-dimensional images and can be explored in a 3D space (Maiwald et al., 2018), and compare the historical and current states of Cultural Heritage objects (Al Khalil a, P. Grussenmeyer b 2019). The photogrammetric technique, in fact, offers the opportunity not only to reconstruct accurately the geometry of the object but also to collect important data such as texture, materials, colour, damages, information necessary for the knowledge of its structure. In some particular conditions, without appropriate documentation, the historical images are the only way to study and reconstruct the past (Bitelli et al 2007).

Classical analytic photogrammetry tackled the issue of processing historical images since their birth. Recently, the advancement of new technologies is supporting the calibration and orientation steps in the photogrammetric pipeline making them automatic. However, although the use of Structure-from-Motion

(SfM), the position and orientation of the camera, cannot always be automatically estimated with these algorithms implemented on historical photographs.

The calibration of these images is the key step and some critical factors could lead to a failure of the processing.

In the following Figure 1, 2 and 3 The Valentino Castle, the seat of Politecnico di Torino, has been chosen as an example of some of the issues following mentioned.

Issues related to archive storage

The first problem is the difficulty in finding suitable material for photogrammetric processing with a good conservation stage. Low archival quality due to improper transport or storage procedures of the film (humidity, temperature, etc.) together with inaccurate processing of original films or hardcopies in field laboratories could cause issues of digitization.

Film grains, dust particles and fingerprints may appear on the images and may also be visible on the digital copy. In most cases, there is no information about the scanning process, and if the scan information (sensor, resolution, dynamic range, working range, precision, filter) is not available, all metric data could be damaged (Maiwald, 2019).

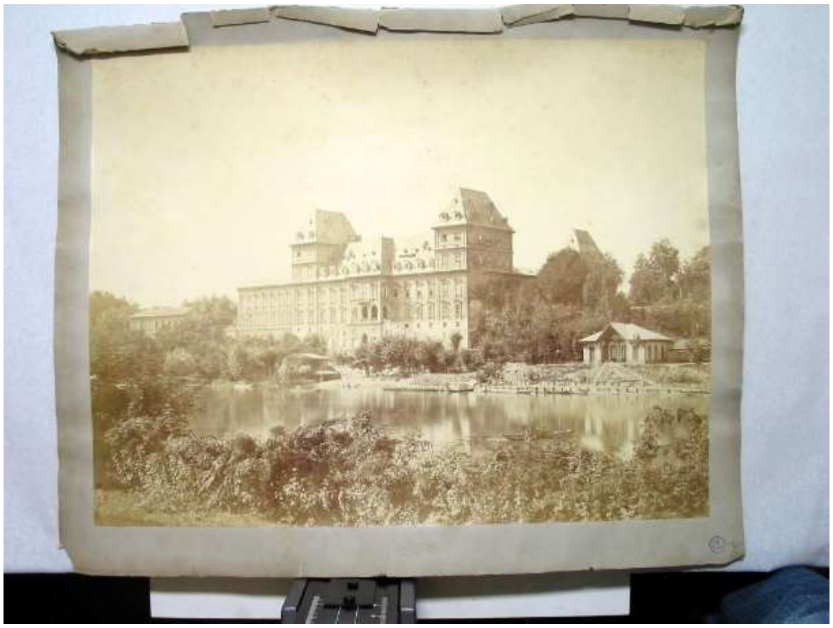


Figure 1. Historical photographs of the Valentino Castle, from the archive of GAM - Galleria Civica d'Arte Moderna e Contemporanea, in which some archive issues (bad conservation stage due to humidity) occur.

The consequence is a low quality of images in terms of resolution and radiometry (haze, image darkness, non-uniform luminosity); the lack of information about the images (acquisition period, method and location); the difficulty in finding accurate data like constraints or control points (Bitelli et al., 2007; Zawieska et al., 2017; Maiwald et al., 2017).

Issues related to the aim of the acquisition

Another challenge is relating to the fact that historical photographs and film footage stored in archives were not taken to be used for metric documentation and 3D reconstruction purposes. In the majority of cases, they consist in the press and personal memories for what concerns photographs; while for film footage they consist of movies, amateur videos or cinematographic reports. For this reason, images could be incomplete or could present occlusions due to persons or car in front of the object to reconstruct.

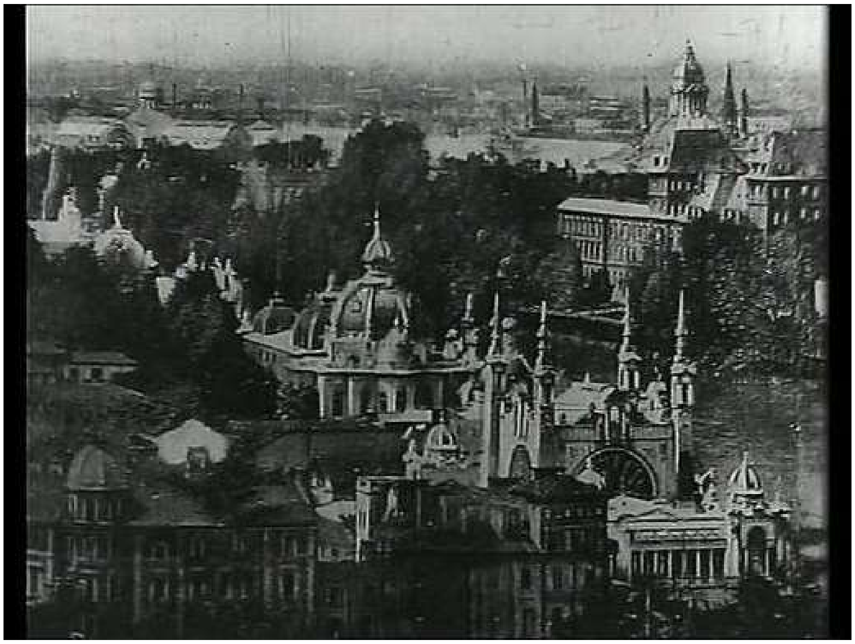


Figure 2. Historical film footage of the Valentino Castle, from the I-Media-Cities platform and the video archive of Cinema Museum of Turin, in which some aim of acquisition issues (documentary) and technical features issues (low resolution quality and lack of film information) occur.

Issue related to technical features

Another issue is that camera parameters and film information are often missing. The inaccuracy or total lack of meta-information about inner orientation (focal length and coordinates of principal point and fiducial marks when present) and additional (i.e., distortion) parameters are the most important problems because they are necessary to perform the interior and exterior orientation. Moreover, the scanning process cloud affects the individuation of the principal point when only a part of the original analogue image is scanned and it may be at the edge or even completely outside the digital image.

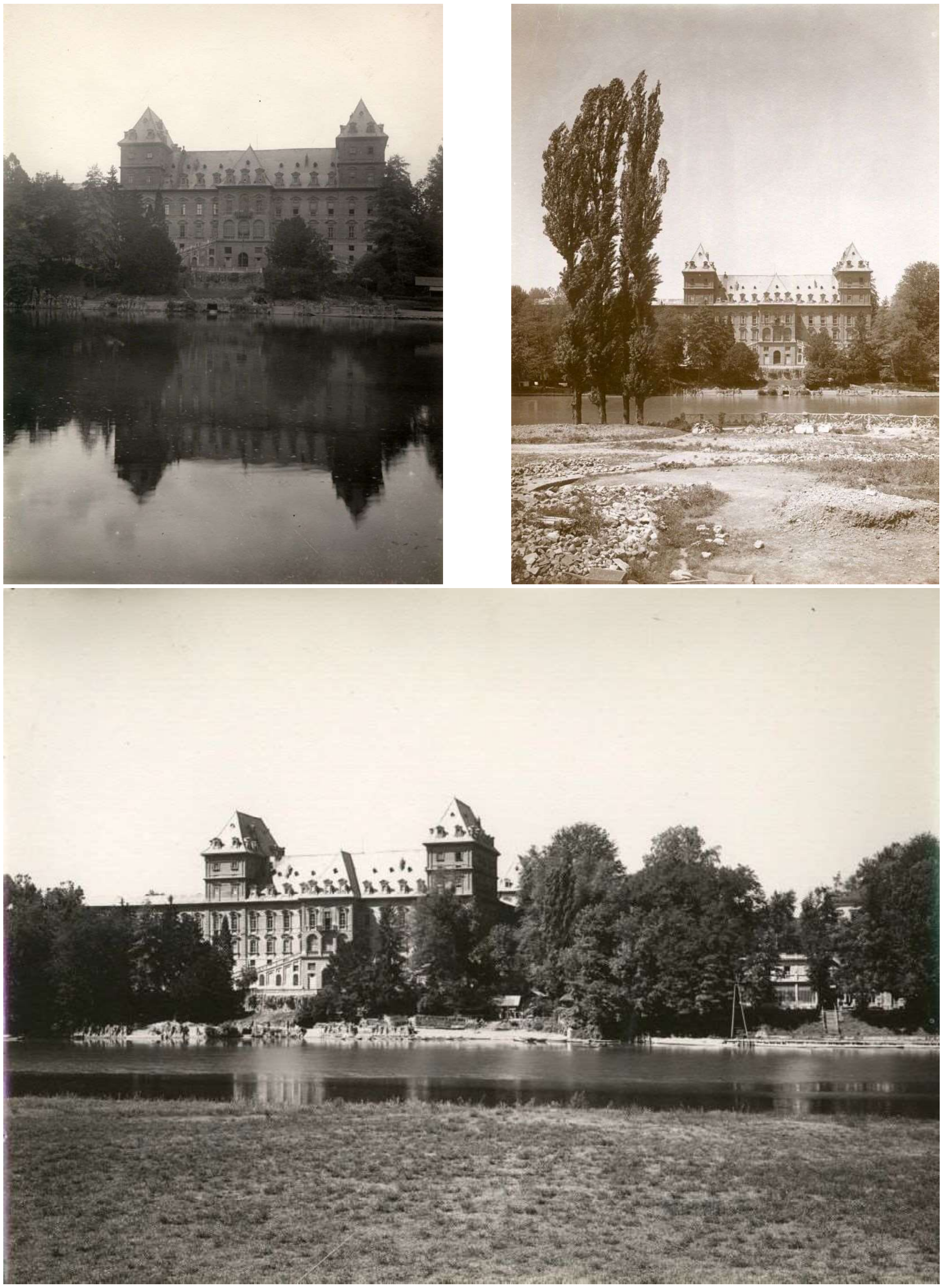


Figure 3. Historical photographs of the Valentino Castle, from the archive of GAM - Galleria Civica d'Arte Moderna e Contemporanea – Fondo Gabino, in which some technical features issues (images of the same object are taken by different types of camera that differ in exposure and shooting) occur.

Problems increase when images of the same object are taken by different types of camera that differ in exposure and shooting and are used in the same reconstruction. The season, the weather, the time of the day in which images were acquired dramatically influence the changes in radiometric characteristics. Moreover, the images may be blurred, noisy, underexposed and overexposed, and different points of light, reflections and shadows may appear in the same scene and interfere with feature extraction during the photogrammetric process. In addition, extreme changes in perspective can occur between images, making 3D reconstruction difficult.

The last issue concerns images acquired in a different historic period in which the building could be affected by changes during time, such as destruction or reconstruction of some parts. These details could be useful during the orientation process, but at the same time could lead to some mistakes in the reconstruction.

2.1.2 Classification of historical images and state of the art

With the awareness of the limitations concerning the processing of historical images highlighted in the previous section, in the following paragraphs a classification of historical images is presented. Moreover, the state-of-art and potentialities of these kinds of primary data are highlighted, focusing on the previous attempts in managing these images.

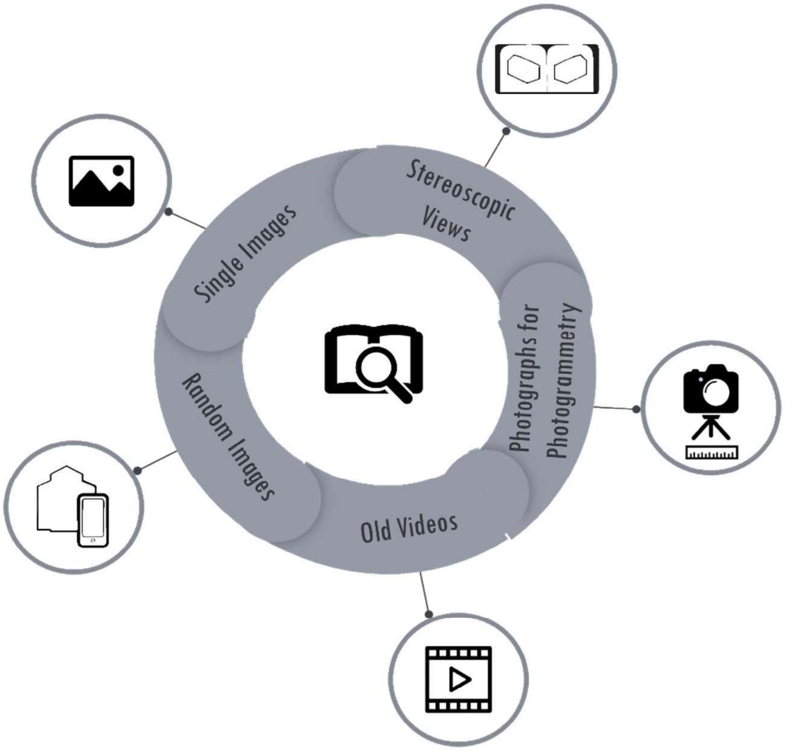


Figure 4. Classification into five main categories of historical images stored in historical archives, according to their aim of acquisition and their availability.

The proposed approach splits historical images into five main categories, as shown in Figure 4, according to the aim of acquisition and the availability in archives: (1) single images; (2) photographs for photogrammetric purposes; (3) random images; (4) stereoscopic views; (5) videos.

Single Images

Recovering metric information from a single image is a classical problem in photogrammetry, 3D reconstruction, computer vision and robotics. In photogrammetry “three-dimensional reconstruction of an object from a single image is only possible if additional geometric information about the object is available. Single image processing is typically applied for and orthophotos, and plane object measurements. The achievable accuracy of object measurement depends primarily on the image scale and the ability to distinguish those features which are to be measured within the image” (Luhmann et al., 2020).

In the field of Cultural Heritage is particularly important in cases in which no more than one view of a monument to document is possible to recover.

In the Cultural Heritage domain, some studies have been carried out for the rectification of historical images that can efficiently provide measurements from a single image using the a priori geometric properties of the object such as linearity, parallelism, perpendicularity, symmetry (van den Heuvel, 1998). The methodology includes the determination of segments and vanishing points and, if possible, the intrinsic calibration of the camera (Bräuer-Burchardt and Voss, 2002). The knowledge of a reference geometry allows estimation of the reference scale. Rectification, for example, is a technique particularly suitable for flat façades (Khalil and Grussenmeyer, 2019) where the low relief (if present) is assumed to produce deviations from the orthogonal projection since some details do not belong to the rectification plan. They are negligible depending on the scale chosen or the use of this rectification. In case of the destruction of the object, sometimes the reference information is provided by the still existing buildings in the neighbourhood (Hemmleb, 1999). If the parameters of the camera model, such as the focal length, are unknown, they can be estimated from the geometric information of the object. The rectification, when possible, estimates eight transformation parameters which are the coefficients of the equations of the general homography. They derive from the collinearity equations in the hypothesis that the object is a perfect plane. The eight parameters of the homography contain functions of all nine internal and external orientation parameters.

Other approaches developed a system to solve the problem of reconstructing the geometry of an object from a single view: recognizing selected objects in uncalibrated images (Bryson et al., 2014); jointly analyzing a collection of images of different objects along with a smaller collection of existing 3D models to reconstruct the scene from a single view (Huang et al., 2015); recovering 3D object from single 2D line drawing in perspective projection reconstruction without knowing camera parameters (Yang and Zhang, 2017); measuring objects images of unknown origin using a Python tool that estimates vanishing points and resects a camera and poses is also proposed (Settergren, 2020).

Recently, with the spread of Artificial Intelligence techniques, this topic is again addressed proposing different solutions. Following some example of applications are reported. Some approaches deal with the problem considering that a depth image from one viewing angle may be associated with an infinite number of possible 3D models. Starting from the ability of the humans to solve such ambiguity, some researchers applied supervised learning combining monocular and stereo cues to estimate the depth map of the single image (Saxena et al., 2008); trained neural networks to acquire 3D shape from single depth view (Yang et al., 2018) and to estimate the 3D object shape from 2.5D sketches (Wu et al., 2017).

In the heritage domain, a study focused on the single photo 3D reconstruction problem for lost cultural objects for which only a few images are remaining using an image-to-voxel translation network (Z-GAN) as a starting point. (Kniaz et al., 2019).

Photographs for Photogrammetry

Starting with the case of many photographs in Meydenbauer's metric image archive, analogue before, and digital later, have been used for documentation purposes. If the photogrammetric acquisition rules have been observed and calibration data or camera position information is still available, a correct reconstruction is possible under certain conditions. Thanks to the availability of the calibration report of the metric camera used, generally, the generation of the 3D model succeeds at a high rate, even if the images processed are old. The success of a photogrammetric reconstruction also depends on the number of images of an object, ideally taken with the same camera (Hemmleb, 1999). In the past, data processing was carried out on digital stereo plotters or workstations; all analogue photos had to be converted into digital ones using a photogrammetric scanner. Today there are still photogrammetric materials, but the technology and processing are completely different. Many SfM software is available and the commercial type is the most used due to their easy use and the acceptable quality in many applications of the results obtained in a fast way.

Several studies (Pavelka et al., 2017; Poloprutskýa et al., 2019) tried to process different types of photogrammetric images with different results. One of the most used software in previous works is Agisoft Metashape, which leads to results only when the historical images have high overlap.

Stereoscopic Views

In stereoscopic imaging, two images are acquired by different camera poses in such a way that the optical axes of the two camera systems are perpendicular to the base vector and almost parallel to each other (Khalil and Grussenmeyer, 2019). Also converging views are suitable in stereoscopic views and the overlapping is guaranteed. However, while classical photogrammetry solves the reconstruction with overlapping starting from 30% between two images, Structure-from-Motion techniques, which are suitable for many converging images, require an overlap of

85% between images on at least three photographs. For this reason, the processing of stereo image pairs fail. Moreover, increasing the percentage of overlapping, the precision quickly decays (photogrammetric suggestion is to never exceed 60%). This has been demonstrated in a recent study in processing old collections of analogue negatives from terrestrial stereoscopic photogrammetry of historical buildings (Rodríguez Miranda and Melón, 2017). There was almost no overlap in the calculation of relative orientation, and in this situation, the SfM algorithms are not capable of automatically resolving the absolute orientation of the photographs. Traditional photogrammetry can fill this gap and provide adequate results.

Random Images

When images are acquired not following the required photogrammetric criteria they are called "random" in this dissertation. This means that they not present overlapping and are often divergent. They are acquired not for documentation purposes and not using a metric but a consumer camera, even if expensive but more suitable to give a high photographic performance. They are mostly scanned from analogue images or they are acquired from various sources such as the internet or books. The main problems with these images are the differences in size and scale, the unknown pixel size and focal length of the camera, and most importantly, the different shooting times; thus, some parts visible in one image are missing in other images and it is impossible to determine the homologous pixels. The light conditions (shadows) are generally very different, which can lead to problems with automatic matching procedures (Gruen et al., 2014) because some details are not displayed correctly and as a result, too few characteristic points are found, which means that the orientation of the image can be estimated inaccurately or not at all. In the automatic subsequent extensive matching process, many wrong point assignments were found, which led to wrong values in the final point cloud. Also, radiation or blur distorted the result (Maiwald et al., 2017)

Various photogrammetric techniques and algorithms are used to solve these problems. Undoubtedly the bundle adjustment presents a lot of advantages (Gruen et al., 2014), but the most popular are structure-from-motion and image-based rendering algorithms that can estimate the camera information from the images themselves using computer vision techniques (Snavely et al., 2008; Schindler and Dallaert, 2012). Data from these studies suggest that integration into the process with current measurement data and images of the same object (Maiwald et al., 2017) provides many reliable control points for the photogrammetric determination of the internal and external orientation parameters of historical photographs (Hanke et al., 2015). Starting from approximate values of unknown parameters and control points, which are acquired today but can also be identified in the past, and assuming that these points have remained unchanged during the elapsed period, it is possible to solve the orientation problem in order to achieve a good convergence of the reconstruction (Bitelli et al., 2007).

Previous researches were involved in the use of these algorithms on unstructured collections of the archive using touristic photos images to 3D model the scene captured (Snavely et al., 2007) and to reconstruct lost heritage (Gruen et al., 2004; Khalil and Grussenmeyer, 2019).

Videos

If no images are available, videos are actually an interesting source for 3D measurement technique. Previous research has defined "Videogrammetry" as a measurement technique based on the principles of "photogrammetry" (Gruen, 1997). The accuracy of a videogrammetric system and the results obtained in various studies was certified as high as a laser scanner acquisition (Gruen, 1997; Herráez et al., 2016). Another advantage is the reduction of the computing load and thus higher efficiency (Sung et al., 2017). For historical videos, the advantage is the possibility to extract images from hundreds of video sequences and to select stereo pairs (Herráez et al., 2016), which can be processed with photogrammetric techniques.

2.1.3 Structure-from-Motion pipeline with open source algorithms

In computer vision, an automatic calibration and reconstruction process is preferred. For this reason, the current state-of-the-art image orientation SfM algorithms and dense image matching (Multiple View Stereo) allows non-experts to obtain 3D point clouds from uncalibrated images without requiring specific settings in the software of elaboration, even if with different geometric and visual quality. Pipelines are generally quite robust and reliable, most of them are capable of processing even large series of unordered images (Stathopoulou et al., 2019). Commercials software, in particular, present many advantages such as the fully automatic process, the easy use and the reliability of the results.

Among commercial software, Metashape (Agisoft, 2020), Pix4D and many others, are the most used also in the research domain because offers many interesting features like photogrammetric triangulation, point cloud data, measurements for distances, volumes and areas, 3D model generation, orthophoto and textures and the easy use.

However, the high cost of the software licenses, and above all the fully automated approach is a disadvantage because it does not allow any intervention in the reconstruction process, no information about calibration accuracy, image orientation and final model (Bartoš et al., 2014). Commercial software, in fact, with closed source code not allow customer-specific parameterization, which can often lead to misleading results or the use of black boxes.

Moreover, in the case of historical images, this software fails to capture images with low overlap, poor resolution and missing metadata, and deliver fully satisfactory results in terms of completeness and robustness. Therefore the final point cloud, if obtained, results uncomplete, not dense and with low accuracy.

Nowadays, besides commercial software, a large number of freely photogrammetric software or algorithms that contain image processing routines are available. A typical pipeline starts with image orientation based on feature matches between images and the sparse point cloud triangulation (Structure by Motion - SfM) using incremental and/or global Bundle Adjustment (BA). Next, a dense 3D reconstruction (usually called Multiple View Stereo - MVS) is performed to further densify the sparse point cloud by reconstructing the depth value of almost every corresponding pixel in 3D space. Open source solutions are usually not designed to support 3D scale reconstruction using ground control points (GCP), but rather with a simple Helmert transformation (Stathopoulou et al., 2019).

The use of open-source algorithms allows the control of the quality of the results at each stage of the photogrammetric pipeline and avoids the blind automatisms of commercial software packages. However, a certain level of technical expertise and knowledge is required. The advantage is to offer the possibility to choose different levels of automatization, parameterization and customization of the algorithms at the basis of the pipeline.

Previous studies (Bartoš et al., 2014; Rahaman and Champion, 2019; Stathopoulou et al., 2019) gave an overview of the most used SfM software, both commercial and open-source, with the aim of comparing their workflows and outputs according to different parameters.

A comparison was performed between commercial and free software for the process of archive images considering reconstruction density, point cloud consistency and quality of the 3D mesh model. The software was selected on the basis of their price, platform independence, scalability and output format (Khalil and Grussenmeyer, 2019).

Another evaluation considered accuracy, ease of use and installation, and the required processing time of open source software and their performance in reconstruction was compared with the results obtained from commercial software to assess the average deviation of their produced point clouds in order to choose the best one according to the final use (Rahaman and Champion, 2019).

Three of the available commonly used open-source solutions, COLMAP (Schönberger et al., 2016), OpenMVG+OpenMVS (Moulon et al., 2016) and AliceVision (Moulon et al., 2016; Jancosek et al., 2011)., were evaluated under diverse large scale scenarios with the aim to check algorithm reliability and performances on large and extensive datasets (Stathopoulou et al., 2019).

The above mentioned open source solutions are mainly developed by the computer vision community, are aimed at a broader audience of 3D reconstruction. Their main goal is therefore not metric accuracy, but photorealistic 3D models of any scale and low geometric quality. On the contrary, MicMac3 (Pierrot-Deilligny and Paparoditis, 2006; Rupnik et al., 2017) is a fully photogrammetric open-source pipeline that can handle GCP and camera constraints (e.g. fixed baselines, etc.). MicMac (http://micmac.ensg.eu, 2020) has been developed at the National Institute of Geographic and Forestry Information (IGN) and the National School of

Geographic Sciences (ENSG), and the access is by simplified command line. The use is not so easy for the unskilled operator because of many small processing steps, but the estimation of the camera orientation parameters and the matching algorithms are well defined and stable (Rupnik et al. 2017).

Despite the high control and stability of MicMac, in this thesis COLMAP was chosen as reference software, as explained below in detail.

2.1.4 Solutions proposed: COLMAP

COLMAP (Schönberger and Frahm, 2016) open-source Structure-from-Motion and Multi-View Stereo (MVS) algorithm implementation, developed by ETH of Zurich (https://github.com/colmap/colmap, 2020), is the pipeline chosen as the reference in this work.

This software is designed to create a versatile incremental SfM system for the reconstruction of collections of unordered photographs. The advantage of COLMAP is that the accuracy of the results is significantly improved while increasing efficiency at every stage of incremental reconstruction. Moreover, it allows the setting of a suitable scenario also for video sequences and supports both graphical and command-line interface. These advantages allow a suitable interaction for the purpose of the research.

The steps of the COLMAP SfM sequential processing pipeline for the iterative reconstruction are: 1) Feature detection and extraction, 2) Feature matching and geometric verification, 3) Structure and motion reconstruction.

In the first step, feature detection and extraction find sparse feature points in the image and describes their appearance using a numerical descriptor. For the feature detection, the SIFT algorithm (Lowe, 2004) is implemented with the possibility to choose both CPU and GPU options.

In the second step, feature matching and geometric verification find correspondences between the feature points in different images. A list of settings is available: exhaustive matching, sequential matching, vocabulary tree, spatial matching, transitive matching and custom matching.

The 3D reconstruction step is performed by implementing an incremental SfM from a carefully selected initial image pair and applying a robust algorithm to select the next view, followed by multi-view triangulation. During the bundle adjustment phase, the Ceres solver and global BA are used at each step to improve camera and point estimations and to avoid drift (Schönberger and Frahm, 2016). The multi-view stereo reconstruction is performed based on the framework of (Zheng et al., 2014) using a stereo approach based on probabilistic patches (Schönberger et al., 2016; Stathopoulou et al., 2019).

During the process, the Final Cost values are computed. They represent the average of the reprojection error overall image observations and are expressed in pixel. It means that they describe a measure of dissimilarity that is the unlikelihood that two pixels belong to a unique point in 3D since the correspondence can be ambiguous.

All data processed during the process are stored in a customized database and could be easily managed.

In the next sections, it will be presented how the standard COLMAP pipeline was customized for the purpose of processing historical images.

Other studies (Maiwald et al., 2018; Maiwald, 2019) focused on the experimentation of different matching algorithm to find the one suitable for historical photographs.

In this study, a method to optimize the feature extraction in COLMAP for historical archives images and then the settings of the scenario suitable for the processing videos are presented following.

2.2 Optimizing feature extraction in COLMAP

In the previous Chapter 1, the importance of metric quality certification of the 3D model obtained from a photogrammetric process was highlighted. Moreover, the use of open-source software proved to be suitable for processing historical images since it allows the adaptation of specific algorithm in the photogrammetric pipeline.

Among the issues observed in the processing of materials from historical archives, such as photographs and videos, one major problem is represented by the fact that the resulting point clouds present low density. This limitation, due to the lack of necessary information in the photogrammetric reconstruction, dramatically affects the evaluation of the metric quality of the resulted point clouds.

Point cloud generation and 3D modelling are two different things. Point clouds are not a 3D model but the starting point of the reconstruction. In modelling, many simplifications are introduced so that the quality can only decrease. By reasoning about points, when possible, a real idea of the accuracy of the measurement method can be offered. When it is not possible, the errors found depend not only on the measurement phase but also on the modelling phase.

A very common practice to evaluate the metric quality of the 3D reconstruction process in photogrammetry is the point clouds comparison. The metric comparison between point clouds with different density is limited by the fact that the 3D model is constituted by incomplete parts. Consequently, assessing the quality of point clouds is, therefore, a challenging problem, since this 3D representation format is unstructured (Javaheri et al., 2017).

To solve this problem, several approaches have been proposed in previous studies: outliers filtering (Hu et al., 2019) and noise smoothing (Wang et al, 2013); automatic filtering procedure based on some geometric features computed on the sparse point cloud created within the bundle adjustment phase (Farella et al., 2019). These studies have focused on finding automatic solutions to the problem.

However, when historical archive material is processed, the difficulty of comparing sparse point clouds lies in the fact that there is no direct correspondence between each point in the two clouds (Tazir et al., 2018). It represents a major problem when it is necessary to scale the model for its metric quality assessment. If point clouds from a recent survey are available, the comparison between this point cloud and the one resulting from the photogrammetric processing of historical images could fail because the few points of the sparse cloud do not match those of the dense point cloud. The situation is even more difficult if there is no current survey, but only historical drawings and archive projects from which the distances to scale the model can be extrapolated.

Developing new ways to improve the performance of open-source Structure-from-Motion algorithms for assessing the metric quality when the resulting point clouds present low density or when dense point clouds are not available is great of interest not only for the Cultural Heritage field but also in other applications.

For example, point clouds, coming from different primary data and/or techniques (e.g. the ones coming out from photogrammetric survey and the ones coming out from a laser scanning process), could vary greatly in their point densities and their accuracies. This is due to the intrinsic characteristics of the instruments, the sensor size and the distance between sensor and object (Bracci et al, 2018). Especially in photogrammetry working with sparse point cloud happens very often, for example when the surface of the object or scene is difficult to reconstruct because is shiny (Delis et al., 2017), texture-less (Hafeez et al., 2017) or curved (Wong & Chan, 2010). The result of point clouds is also affected by bad illumination conditions (Girardeau-Montaut et al., 2005), the thickness of the object and its transparency. Above all, the way in which the data is acquired can cause noisy results and blunders especially when different platforms or low-cost sensors (Byrne et al., 2017) are employed, due to scale and illumination changes or quality and quantity of single sources (Farella et al., 2019).

This section investigates how the performance of open-source SfM algorithms can be improved in order to guarantee the presence of strategic feature points in the resulting point cloud, even if it is sparse. To achieve this, a photogrammetric workflow is proposed to process historical images (Figure 5). The first part of the workflow introduces a method that allows the manual selection of feature points during the photogrammetric process. The second part evaluates the metric quality of the reconstruction on the basis of a comparison with a point cloud that has a different density from the sparse point cloud. This procedure could be also useful in case of a lack of point clouds to be compared: the presence of some specific known points, selected by the human operator, will allow the correct scaling of the obtained point cloud.

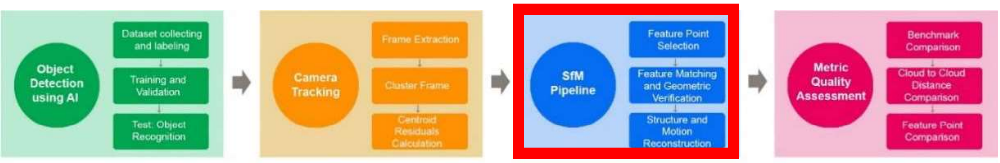

Object Detection using AI
Dataset collecting and labeling
Training and Validation
Test: Object Recognition
Camera Tracking
Frame Extraction
Cluster Frame
Centroid Residuals Calculation
SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction
Metric Quality Assessment
Benchmark Comparison
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 5. The pipeline described in this section 2.2 corresponds to the third step highlighted in red.

2.2.1 Proposed workflow

As introduced in Section 2.1.4, the photogrammetric pipeline chosen as a reference in this workflow is the COLMAP (Schönberger et al., 2016) opensource Structure-from-Motion and Multi-View Stereo (MVS) algorithm implementation, developed by ETH of Zurich, (COLMAP, Johannes L. Schoenberger, 2019).

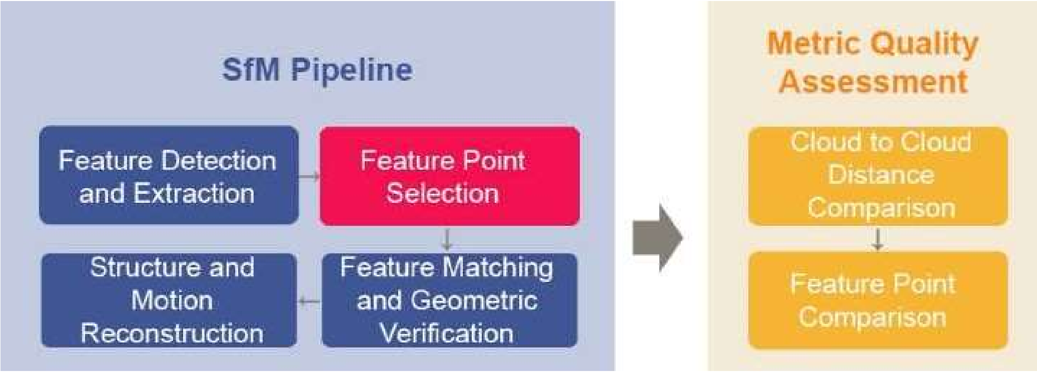

SfM Pipeline
Feature Detection and Extraction
Feature Point Selection
Structure and Motion Reconstruction
Feature Matching and Geometric Verification
Metric Quality Assessment
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 6. Flowchart of the proposed workflow in which a step (in red) of "Feature point selection" was added to the standard SfM pipeline in COLMAP.

Two main blocks compose the proposed workflow. The first one is the standard photogrammetric pipeline in which an additional step of "Feature point selection" (highlighted in red in Figure 6) was added, after the "Feature detection and extraction" phase in order to manually select the tie points to use during the subsequent "Feature matching and geometric verification" step.

The second block consists of evaluating the metric quality of the results obtained from the previous photogrammetric process. In order to reach this objective, firstly the cloud-to-cloud distance and then the Residuals were estimated.

2.2.2 Feature point selection

The algorithm to detect and extract new feature from images in COLMAP is sketched in Figure 7 and detailed below:

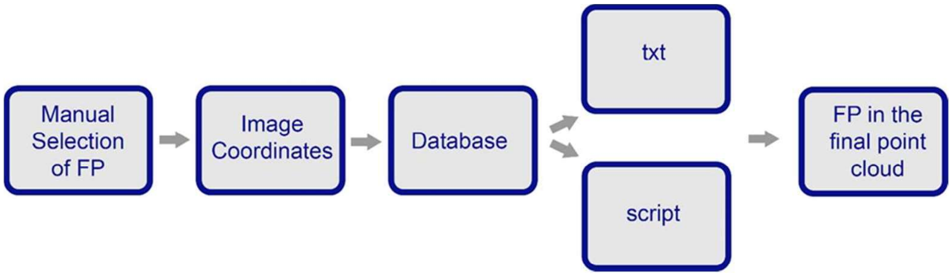


Figure 7. Workflow of the step of the Feature point selection algorithm.

- *Manual selection of Feature Point*: with the standard "Feature detection and extraction" step, COLMAP automatically detects key points in the images, but it could occur that some important radiometric corner in the image, that appear also in other images, are missing. Introducing this step, it is possible to manual detect the feature point of interest and to extract their 3D coordinates. The *image coordinates* of the searched point were measured with the WebPlotDigitizer tool (https://automeris.io/WebPlotDigitizer, May 2020), as shown in Figure 8, and inserted in the software choosing between two different methods, as explained in the following stage.

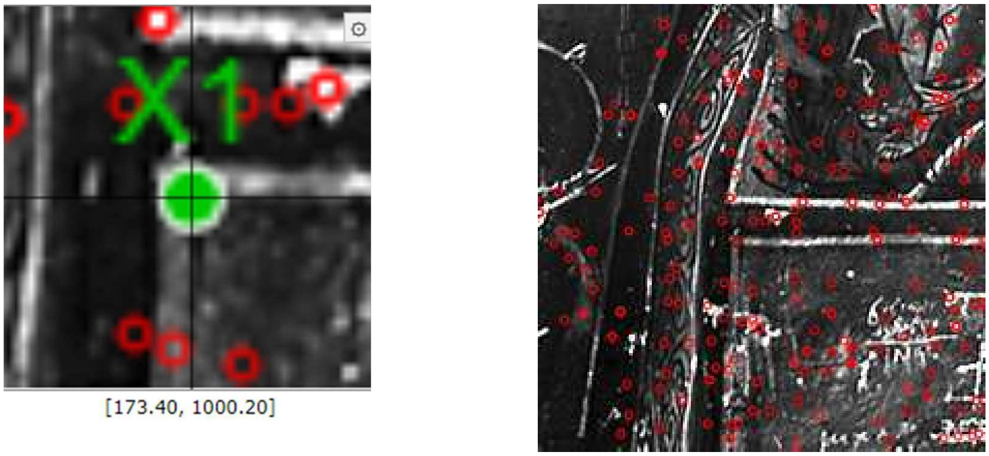


Figure 8. An example of image coordinates measured with WebPlotDigitizer tool and inserted in the software for the manual detection of Feature Points.

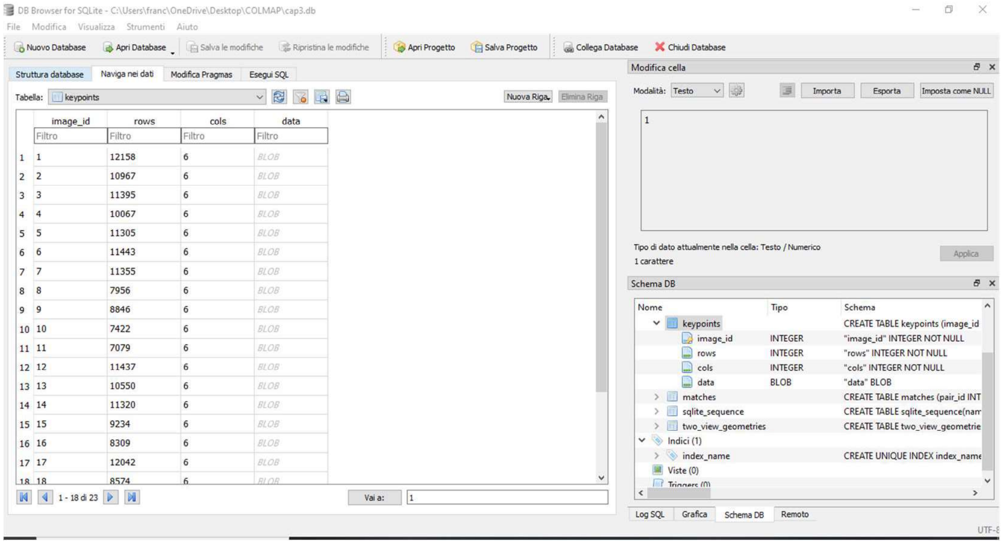


Tabella: keypoints

	image_id	rows	cols	data
	Filtro	Filtro	Filtro	Filtro
1	1	12158	6	BLOB
2	2	10967	6	BLOB
3	3	11395	6	BLOB
4	4	10067	6	BLOB
5	5	11305	6	BLOB
6	6	11443	6	BLOB
7	7	11355	6	BLOB
8	8	7956	6	BLOB
9	9	8846	6	BLOB
10	10	7422	6	BLOB
11	11	7079	6	BLOB
12	12	11437	6	BLOB
13	13	10550	6	BLOB
14	14	11320	6	BLOB
15	15	9234	6	BLOB
16	16	8309	6	BLOB
17	17	12042	6	BLOB
18	18	8574	6	BLOB

1 - 18 di 23

Vai a: 1

Modalità: Testo

1

Tipo di dato attualmente nella cella: Testo / Numerico
1 carattere

Nome	Tipo	Schema
keypoints		CREATE TABLE keypoints (image_id
image_id	INTEGER	"image_id" INTEGER NOT NULL
rows	INTEGER	"rows" INTEGER NOT NULL
cols	INTEGER	"cols" INTEGER NOT NULL
data	BLOB	"data" BLOB
matches		CREATE TABLE matches (pair_id INT
sqlite_sequence		CREATE TABLE sqlite_sequence(nam
two_view_geometries		CREATE TABLE two_view_geometrie
Indici (1)		
index_name		CREATE UNIQUE INDEX index_name
Viste (0)		
Triggers (0)		

Figure 9. An example of the database management in COLMAP in which key points are stored as row-major float32 binary blobs.

- *Database*: data processed in COLMAP are stored in a customized database that could be easily managed. The previously detected key points are stored as row-major float32 binary blobs, (a binary large object that is a collection of binary data stored as a value in the database) where the first two columns are the x and y image coordinates in pixel.

o The first way is useful to insert new coordinates of Feature Points not already detected by the automatic algorithm and that for this reason are not present in the database. Creating a text file in which the image coordinates (x, y) expressed in pixel and the scale and orientation information are indicated, writing one line per feature, it is possible to import known feature (e.g. single points) in the database and use them in the matching stage.

o The second method is to query the database with SQLite to choose the best key points to use for the matching and the incremental reconstruction within all the tie points automatically extracted from the algorithm.

- *Results*: this assisted processing, in both ways, allows the choice and the filter of highlight points such as corners and outline feature. The matching algorithm searches for the selected feature point in each image to estimate the equipolar line in the other images. The result of the matching is a point cloud that, even if sparse, contains the corresponding point to the dense point cloud to compare.

2.2.3 Metric Quality Assessment: Point Cloud Comparison

The metric quality assessment of the photogrammetric reconstruction was added as the last step in the workflow.

This evaluation consists of two main procedures: a global comparison between a sparse point cloud and a dense point cloud and a punctual comparison between selected specific feature points in the two point clouds.

The first comparison was performed using the CloudCompare software (https://www.danielgm.net/cc/, May 2020). This open-source software allows the comparison of point clouds by estimating their distances using the Multiscale Cloud Model Comparison (M3C2) plug-in, which uses the normal directions of one of the two surfaces to calculate local distances and provides estimations of the confidence intervals for each measurement (Lague et al., 2013). Generally this method is a good solution when the two point clouds have the same density. However, in the case analysed in this dissertation, the advantage of this algorithm is that avoids the problems raised by the low density of the point cloud resulting from the photogrammetric process and rather performs a direct comparison of the two point clouds. In fact, thanks to the selection of the feature points in the previous step, the problem of the lack of points in some parts of the point cloud is avoided because the presence of these points in the point cloud is guaranteed. For each point of the sparse cloud, a closer point can be defined in the dense cloud and the algorithm can estimate the surface change as the distance between the two points.

The second comparison was carried out by first scaling the 3D model obtained by the photogrammetric procedure on the basis of the feature points of the dense point cloud, which are identical to the feature points previously selected in the photogrammetric process. Finally, the estimation of the Residuals between the coordinates of the feature points in the two point clouds concludes the metric evaluation.

2.3 A benchmark for historical film footage to assess the metric quality of the photogrammetric reconstruction

2.3.1 Toward the need of a new video benchmark

The previous section focused on the processing of historical images boosting the photogrammetric pipeline with new steps in order to obtain point clouds comparable even if a high density is not present.

For what concerns the processing of video frame and in particular historical film footage, this method is not sufficient to evaluate the metric quality of the models. As highlighted in section 2.1.1, determining the metric precision of the film footage processing is a challenge since the main problem is the way in which the video was shot. The motion of the camera used to shot the film, in fact, dramatically influences the possibility or not to process these data. If it not has created convergent views, the processing failed for the reasons related to the photogrammetric acquisition 3x3 rules (Waldhäusl et al., 2013).

However, as deeply explained in the next paragraphs, specific types of camera motions are selected and considered somehow appropriate for the photogrammetric processing: Up/Down Motion-Tilting, Left/Right Motion-Trucking and Rolling Motion-Panning.

To show the potentialities of the method proposed, this section examines the maximum metric accuracy reachable while implementing photogrammetric workflow on videos shot with these fixed camera motions.

In order to evaluate the metric quality achieved by processing historical film footage with photogrammetric techniques, a benchmark was created on a new video dataset with the aim of reproducing the camera motions in which old videos were shot. The methodology was tested by acquiring videos on Valentino Castle in Turin, a UNESCO World Heritage Site, and processing them with open source algorithms using specific settings in COLMAP.

This is the first benchmark based on video acquisition and processing in relation to film footage.

In fact, despite the importance of this topic, the researchers have not dealt with it in great detail, and previous studies have not addressed the determination of the metric quality assessment of the results of photogrammetric processing of historical film material.

For this reason, the creation of a new benchmark is necessary to achieve the task of evaluating data, sensors and algorithms since it allows the comparison of the results obtained with a univocal approach. Representing the maximum metric quality achievable in this case, it can be used as a point of reference against which results can be compared to have an idea of the quality of the processing performance.

In the photogrammetric and remote sensing field existing benchmarks evaluated sensors, algorithms and methods for data processing (Bakuła et al., 2019). Some past benchmarking activities were promoted by association such as ISPRS with a different aim: evaluating urban object detection and 3D building reconstruction based on airborne image and laser scanner data (Rottensteiner et al., 2012); creating datasets for multi-platform photogrammetry for the orientation of oblique airborne image sets (Nex et al., 2015; Gerke et al., 2016); comparing of indoor modelling methods (Khoshelham et al., 2017); proposing new dataset for multi-view stereo processing (Schöeps et al., 2017); segmenting UAV videos (Ying Yang & Yilmaz, 2018); assessing the performance of the entire image-based pipeline for 3D urban reconstruction and 3D data classification (Özdemir et al., 2019).

For what concerns historical archival material, a benchmark was proposed to orienting historical photographs, experimenting with different feature matching algorithms (Maiwald, 2019).

However, differently from historical photographs, historical film footage needs to consider also the motions of the camera. This represents the great innovation introduced in this new benchmark here proposed and it will be useful also in other applications.

Recently, in fact, the use of video for the documentation of Cultural Heritage sites has become widespread thanks to the increased quality (both radiometric and geometric) and the number of video streams. Of course new surveying sensors and smartphones have encouraged the diffusion of easy ways to capture video sequences and the consequence is the development of methods to derive 3D data from video for different purposes, not only in the heritage field (industrial, computer vision, UAV, for example).

Most of the research on video sequences have been used for recordings with ad hoc cameras, and dense 3D reconstruction from a video has been proposed to obtain an accurate representation of the scene (Pavoni et al., 2016). Aerial video sequences have the same disadvantages as historical films, such as low resolution, blur-motion effects and redundant video images, and can therefore be compared. Previous studies have investigated the possibility of using video images for 3D modelling with commercial software for processing data with Structure-from-Motion (Cusicanqui et al., 2018). However, as already explained, in the case of historical films, the full automation of the software packages leads to no results and there is a real need to control every step of the photogrammetric workflow.

To achieve a successful 3D reconstruction, certainly the overlapping between subsequent frames, the scale of the image, the viewing angle and the baseline are important factors to consider. Baselines have a particularly key role since if too narrow, they are not optimal for triangulation of tie points, while if too wide the matching of detected keypoints resulted difficult to perform. However, if these factors are taking into account, a videogrammetric approach can provide 3D results comparable to those of a photogrammetric solution based on images from a reflex camera (Torresani and Remondino, 2019).

Considering the important role of the respect of the photogrammetric rules required to obtain high metric quality results, in the next paragraphs the identification of the type of camera motions used in the historical film footage and the state of the art in processing them is presented. Then the acquisition of the dataset and photogrammetric workflow used for processing of video frame is described and finally the criteria used for the metric quality assessment and the results are reported in the last paragraph.

2.3.2 Camera motions analysis and related works

In general, in historical film footage, it is very rare to find camera motions taken from multiple points of view of the same object that create convergent views. If they are available, the application of the bundle adjustment method allows the computation of all camera parameters and 3D object coordinates as well as the compensation of the systematic errors. Instead, it is much more common to find the following types of camera motions (also shown in Figure 10):

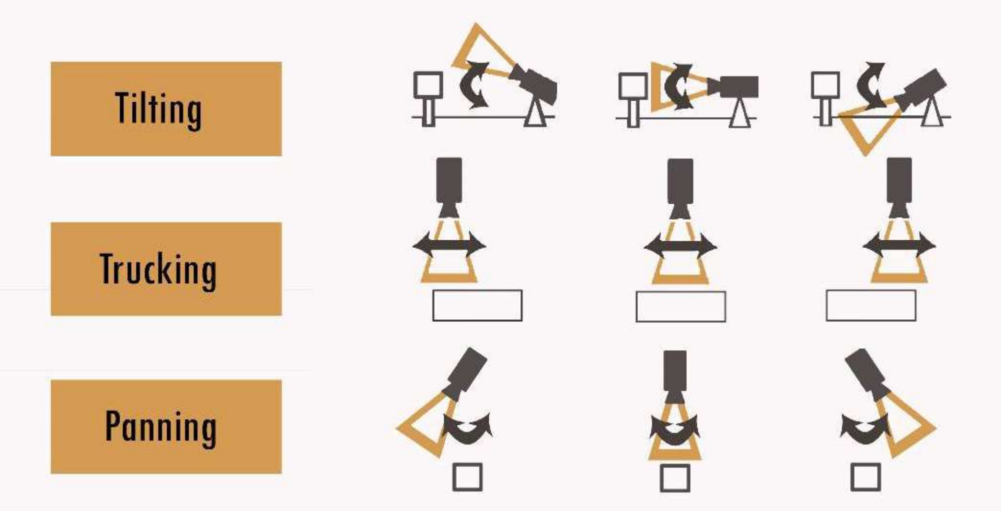


Figure 10. Scheme of the three types of camera motions: Up/Down Motion-Tilting, Left/Right Motion-Trucking and Rolling Motion-Panning.

1) Tilting(Up/Down Motion): camera positioned in a fixed position and that takes the object by scrolling from top to bottom (or vice versa) in a vertical plane.

2) Trucking (Left/Right Motion): camera in motion along with a fixed point and that takes the object by scrolling from right to left (or vice versa)

3) Panning (Rolling Motion): camera positioned in a fixed position and that takes the object by horizontally pivoting from right to left (or vice versa) on a central axis.

In these kinds of sequences acquisition, extracting 3D information is limited not only by the low quality of the frames and the lack of information about camera parameters, but also the noise due to the oscillation of the camera and the small translation. The consequence is that the baseline between adjacent frames is absent or very small, for this reason the bundle adjustment and perspective models could fail for the continuous changes of the internal parameters and the collinearity equations may be ill-conditioned or the rays cannot correctly intersect (Remondino, 2003).

Calibration and reconstruction accuracy increases with the convergence of the images used and with the ratio between the Baseline and the Depth (B/D).

Acquiring converging images can generate a high correlation between system parameters and lead to instability in the estimation of minimum squares. The ideal would be to acquire images at different distances from the object to complete the different correlations between the unknown parameters to be calculated within the bundle. However, these conditions are not always respected due to the movements of the camera.

The following part presents the feasibility and limitations of processing film footage shot with these kinds of camera motions and the analysis of the state of the art for each of them.

Tilting

In the case of a camera that is positioned at a fixed point and captures the object from top to bottom (or vice versa), the baseline is absent or almost null between adjacent images. This could cause problems in the processing because the frames have too high overlap.

There is an absence of variation in the horizontal direction x and a marked variation along y called vertical parallax. The object is however taken from different positions and it is possible to estimate the distance of the camera from the point in a 3D space, because the taken centre moves between frames. Frames filmed with high inclination are to be discarded as they will result in divergent views compared to those filmed with small tilting movements.

In a previous study, photogrammetric analysis of monocular video sequences without the typical photogrammetric information needed to retrieve camera parameters and generate 3D models was investigated. After a series of tests with different data sets it was shown that image orientation and calibration parameters could be successfully determined by knowing the size of certain objects in the scene, the pixel size and a perspective bundle adjustment (Remondino, 2004).

Trucking

In the case of the camera positioned in front of the object and that takes it in motion by scrolling from right to left (or vice versa) there is certainly overlap between adjacent frames. However, the overlap could be not so high and similar to noise due to the instability of the camera.

Previous works treated the topic of how to obtain the camera parameters of orientation and calibration from monocular sequences comparing both the perspective and the projective approaches (Remondino, 2003). Since these sequences were acquired using zooming cameras and with small translations, the projective geometry resulted strongly effective in these cases. The perspective collinearity model, which is highly stable but requires stable optics, is simplified into a scaled orthographic projection that is able to deal with variable focal length and small horizontal translation (Remondino, 2004).

Panning

In the case of a camera that is positioned at a fixed point and captures the object by rotating from right to left (or vice versa), there is no baseline and the classical bundle method cannot solve the adjustment. A previous study has shown that the perspective camera model based on the classical bundle method can be used to calibrate rotating cameras that do not produce cocentric images. Alternatively, a simplified camera model can be used, which simply links the image matches to a rotation matrix. The results obtained with the existing videos do not correspond to the usual photogrammetric accuracy, mainly due to the very low image quality (Remondino and Börlin, 2004).

Therefore, this case could be related to the spherical photogrammetry theorized by Fangi (2007) and subsequent studies. This is an analytical approach that works with a series of images taken from a single point of view to produce a spherical panorama. It is obtained by assembling several images, which are then projected onto a virtual sphere and then mapped onto an equirectangular projection plane using commercial software. If several panoramas of the same scene taken from different angles are available, it is possible to obtain an adequate orientation and a 3D reconstruction of the scene (Barazzetti et al, 2010; Pisa et al, 2011). In these cases it has been shown that it is possible to obtain a good metric content with an average value for the error modulus between about 0.03 m and 0.015 m (D'Annibale et al., 2011).

2.3.3 Acquisition and processing of video dataset for the benchmark

Despite certain limitations suggested by previous studies, processing film footage according to the three types of camera movements identified has been observed as feasible, adopting different solutions. Motivated by the lack of existing benchmarks related to the architecture field, a new dataset is presented with the aim of reproducing the camera motions used to shot architecture and monuments in historical film footage. In particular the tilting and trucking camera motions have been highlighted as the more common for this kind of acquisition. In this evaluation, only these two cases were selected to be analysed in this benchmark, while the case of rolling motion refers to spherical photogrammetry, with the limitation that only a larger number of panoramas can achieve good metric quality, as previously shown.

Video Acquisition

The Valentino Castle in Turin was chosen as a case study and it has been used to evaluate the effectiveness of the experimental methodology presented here. Data for this study were collected by the author under optimal conditions with a calibrated full-frame camera, CANON EOS 5DS R, with a fixed focal length of 20 mm and known settings (focus, aperture, exposure). The following Table 1 shows the specifications of the camera.

Table 1. Camera specifications.

Effective megapixels	50.60
Sensor size	36 x 24 mm
Sensor type	CMOS
Sensor resolution	8712 x 5808
Max. video resolution	1920x1080 (30p/25p/24p)
Focal length	20 mm

Nowadays, most cameras and video devices require a sensor with low power consumption and price and small size. For this reason, the conventional CCD sensors have been replaced by the CMOS sensors, even in expensive cameras such as CANON EOS 5DS R used in this work. One of the disadvantages of this change concerning the video acquisition is related to the rolling shutter. While the CCD sensors present a global shutter where all pixels are reset simultaneously and are able to collect the light in the same time interval, the CMOS sensors have a rolling shutter, in which every row is read and reset in sequence acquiring the image row by row (Forssén and Ringaby, 2010).

Classical Structure from Motion algorithms modelled on the global shutter could fail if applied to rolling shutter because they assume that each image is captured at a single time instance and not row by row. It can seriously affect their performance leading to geometric distortions if the camera or the scene is not static (Hedborg et al., 2013).

Previous studies dealt with the problem of the relation between SfM and rolling shutter proposing methods to model camera motion during image exposure trying to solve the problem interpolating position and orientation (Hedborg, 2012); to rectify video sequences (Forssén and Ringaby, 2010); to assess the difference in metric quality comparing traditional non rolling shutter camera model with a proposed rolling shutter one, demonstrating that the distortions due to rolling shutter increase with the use of a fast moving or vibrating camera limiting the accuracy of photogrammetric reconstruction (Vautherin et al., 2016).

In the case of Cultural Heritage asset, fortunately, most scenes are static. The rolling shutter sensors work without problem because the relationship between the scene and the camera remains unchanged. Problems only occur in dynamic scenes when the object or camera is moving fast and the readout speed of the image sensor is too slow. Under these circumstances, rolling shutter not only reduces the visual quality of the film, but can make 3D reconstruction from the video not feasible, since one image provides multiple points in time in the same frame and each row of the image has its projection parameters (Verhoeven, 2016).

The benchmark presented here is related to architecture object that is for this reason fixed and the scene is consequently static. The video was acquired very slowly to guarantee a high overlap between the frame and to avoid and limit distortions due to the rolling motion as far as possible. For this reason, the problems related to rolling motions are neglected for the purposes of the evaluation performed with the benchmark.

Case study and dataset

Data for the benchmark were acquired by shooting video on the Valentino Castle in Turin.

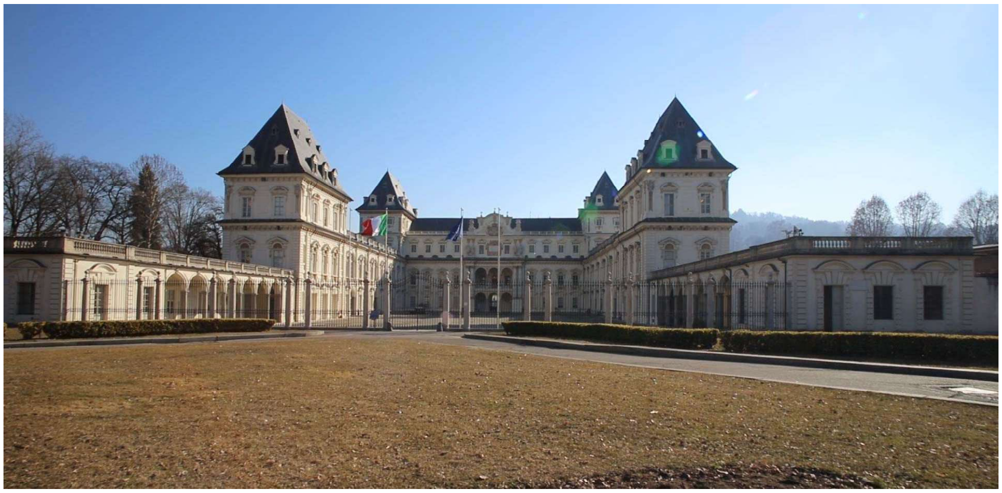


Figure 11. The Valentino Castle, the seat of the Politecnico di Torino.

The building is one of the "Residences of the Royal House of Savoy" and has been on the UNESCO World Heritage List since 1997. The present Valentino Palace of the Savoy dynasty is the result of various design phases that began in the mid-1500s. Following the French *pavillon-système*, the architects Carlo and Amedeo di Castellamonte designed the construction of an imposing building by doubling the existing architectural structure, closed by a pavilion roof and flanked by two tall and slender side towers, connected by terraced porticoes with two new pavilion roofs, towards Turin and connected by a semicircular exedra. Later, in keeping with the eclectic culture, the terraces connecting the two towers were replaced by two large galleries. After extensive modification and restoration work, the castle is now the seat of the Politecnico di Torino (Dameri, 2009).

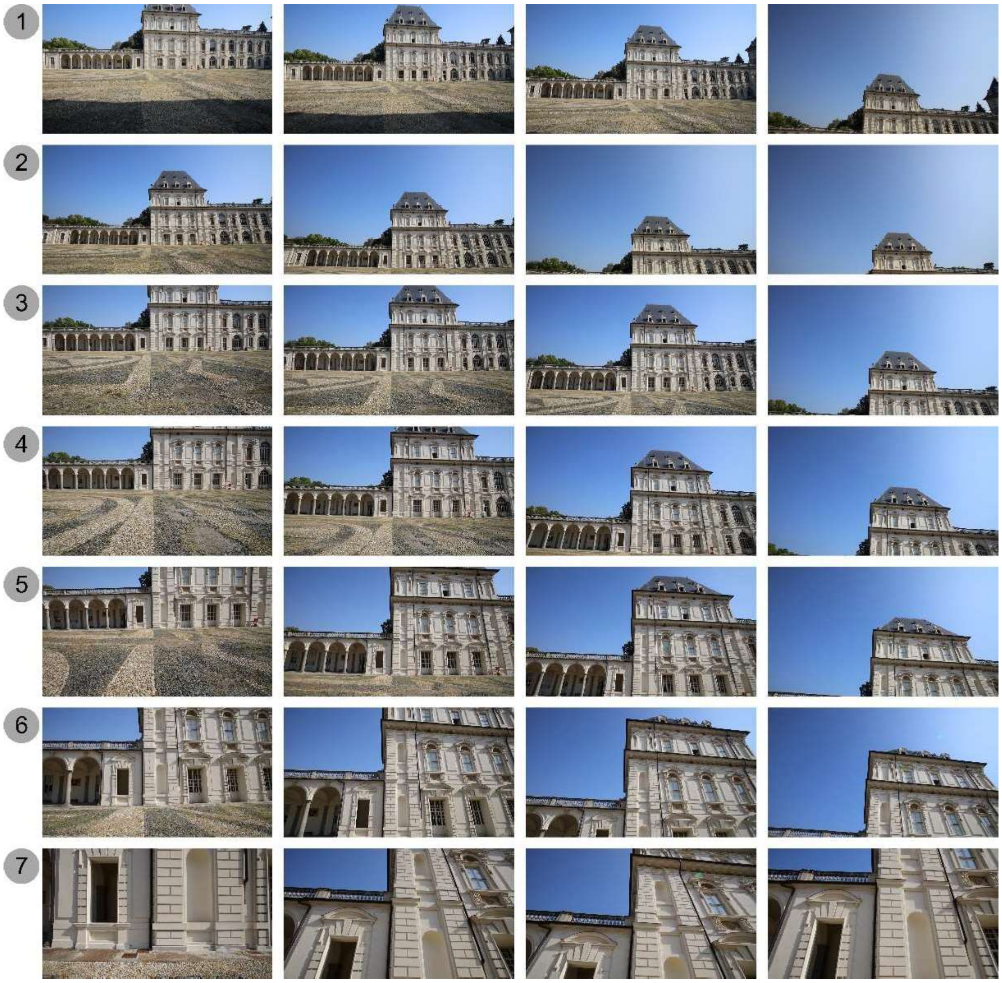


Figure 12. Frames extracted from the video sequences of the left wing of the courtyard of the Valentino Castle shot with tilting camera motion. The sequence 1 was shot at a distance of 43 m, the sequence 2 at 38 m, the sequence 3 at 33 m, the sequence 4 at 28 m, the sequence 5 at 18 m, the sequence 6 at 10 m, the sequence 7 at 5 m.

In order to collect data for the benchmark, the videos were shot recreating the camera motions previously identified as common in film footage and suitable for photogrammetric processing.

Moreover, to analyse the trend of the accuracy values according to the taking distance, the videos were shot at different and fixed distances.

In particular, videos shooting with the tilting camera motion were taken on the left wing of the courtyard of the castle starting with a distance of 43 m from the building and continuing taking it every 5 meters before and 10 then moving closer until 5 m of distance (Figure 12).

Videos shooting with the trucking camera motion were taken on the façade of the castle starting with a distance of 85 m from the building and continuing taking it every 20 meters moving closer until 25 m of distance (Figure 13).

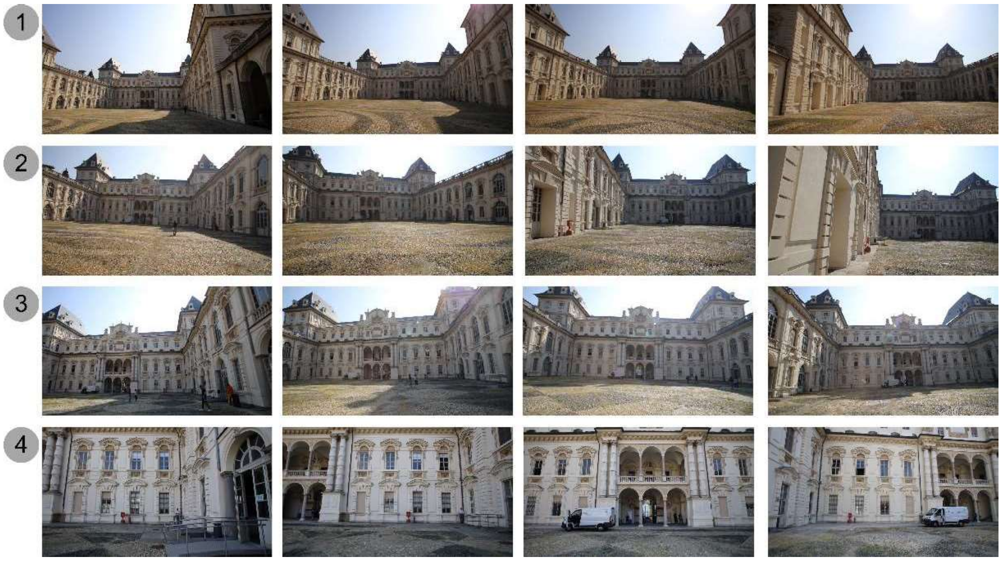


Figure 13. Frames extracted from the video sequences of the façade of the Valentino Castle shot with trucking camera motion. The sequence 1 was shot at a distance of 85 m, the sequence 2 at 65 m, the sequence 3 at 45 m and the sequence 4 at 25 m.

Photogrammetric video processing in COLMAP

Video frames were extracted from videos and processed with the software COLMAP.

The results will become the benchmark for evaluating the quality of historical video processing, and for this reason the settings, the parameters and the workflow followed will be the same for the two different camera motions. In order to follow the same process for the two different cases, the same three steps of the COLMAP SfM sequential processing pipeline for the iterative reconstruction were followed:

1) Feature detection and extraction, 2) Feature matching and geometric verification, 3) Structure and motion reconstruction.

The software allows the definition of different reconstruction scenarios, and in this case the *Video Sequence* is the best way to achieve high accuracy and efficiency, since a video has consecutive frames with a baseline that is too small.

In the first step, *feature detection and extraction* is used to find sparse feature points in the image and describes their appearance using a numerical descriptor. In the best case, as in this case of the benchmark creation, the camera is calibrated, so it is possible to manually specify intrinsic parameters. Generally in the case of historical film footage, only partial or no EXIF information is available, but the software tries to find automatically camera and focal length information. The same camera took multiple pictures with the same lens and settings, so the same information may be shared between all the images. Then the intrinsic camera model must be chosen. In this case, the intrinsic parameters are unknown a priori it is recommended to choose the *Simple Radial Camera Model* that is able to model distortion effects considering the following parameters: f, cx, cy, k1, k2, that is one focal length (f), two coordinates of the principal point (cx, cy) and two radial distortion parameters (k1, k2).

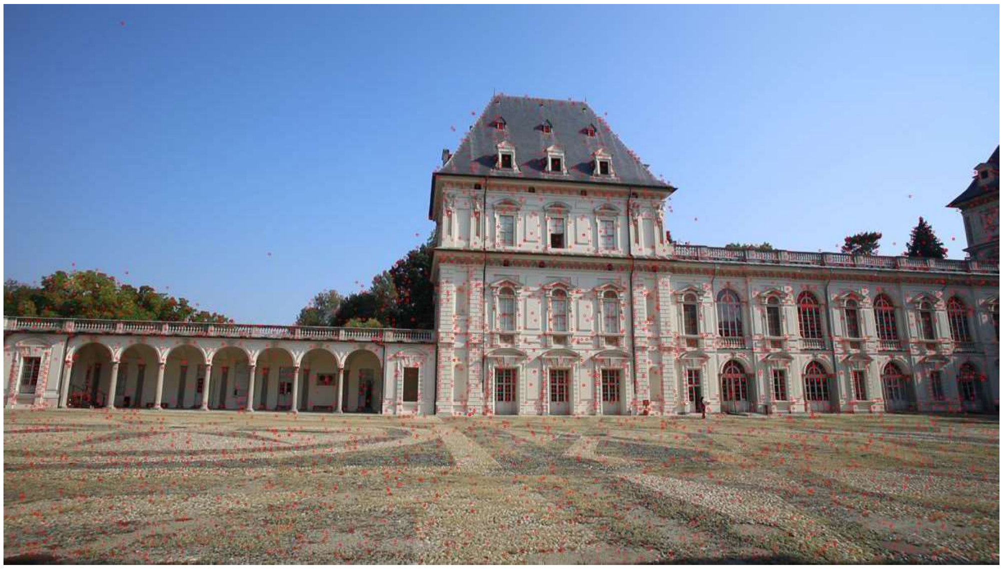


Figure 14. The result of the first step *feature detection and extraction* of the pipeline. The red points are the Feature Points automatically extracted from the software.

In the second step, *feature matching and geometric verification* find correspondences between the feature points in different images. In the case study, it was chosen the *Sequential Matching* mode developed for images acquired in sequential order by a video camera. In this case, consecutive frames have visual overlap and there is no need to match all image pairs exhaustively. For a better reconstruction, the frame rate was reduced, it was increased the overlap and loop detection was enabled.

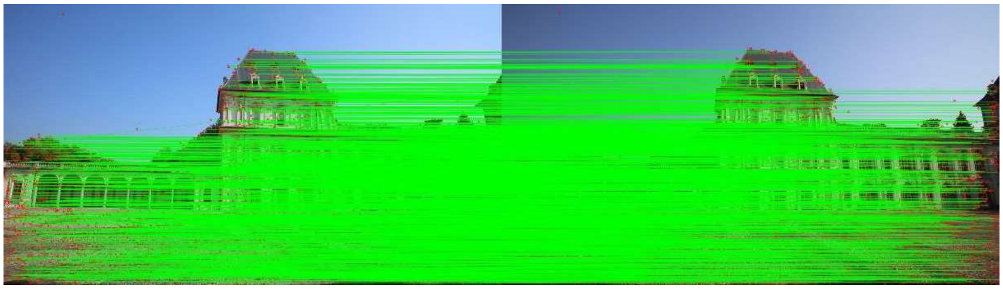
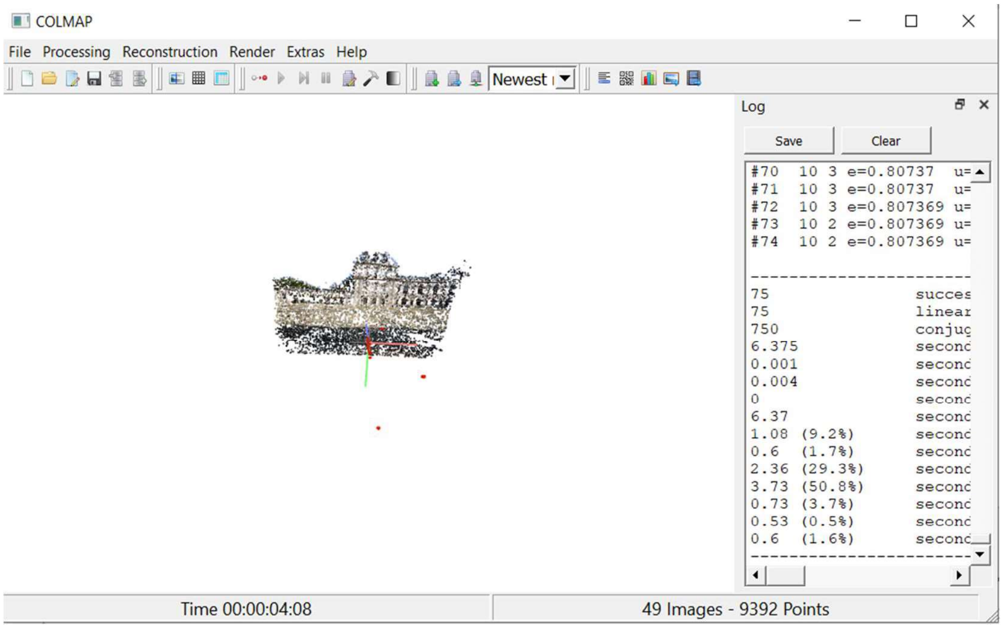


Figure 15. The result of the second step *feature matching and geometric verification* of the pipeline. The green lines are a visualization of the rays that link two homologous points in two consecutive images.

After the matching process, the incremental reconstruction process can begin and the results can be displayed in real time. To get the best results, manual reconstruction was chosen and a file *patch-match.cfg* was written, which contains instructions for the reconstruction. In fact, in these cases, manual selection of the source images with the greatest visual overlap leads to better results, as wider baselines can be obtained by skipping a few neighbours.



```
#70  10 3 e=0.80737  u=
#71  10 3 e=0.80737  u=
#72  10 3 e=0.807369 u=
#73  10 2 e=0.807369 u=
#74  10 2 e=0.807369 u=

---------------------------
75                   succes
75                   linear
750                  conjug
6.375                seconc
0.001                seconc
0.004                seconc
0                    seconc
6.37                 seconc
1.08 (9.2%)          seconc
0.6  (1.7%)          seconc
2.36 (29.3%)         seconc
3.73 (50.8%)         seconc
0.73 (3.7%)          seconc
0.53 (0.5%)          seconc
0.6  (1.6%)          seconc
---------------------------
```

Time 00:00:04:08

49 Images - 9392 Points

Figure 16. The result of the third step of reconstruction. The point cloud results are used for the metric quality assessment.

The processed data are stored in a customized database and can be easily managed. Finally, the results of the analysis and processing of the two types of camera movements are shown in Figure 17 for the tilting sequences and Figure 18 for the trucking sequences, in which some example of feature points detected, feature matching and the final point clouds are reported.

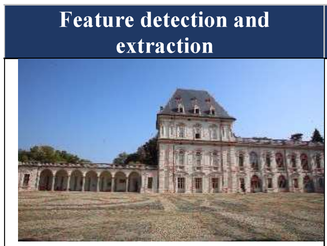
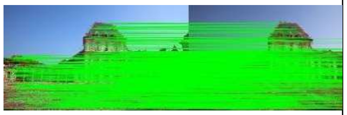
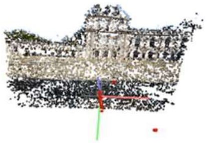
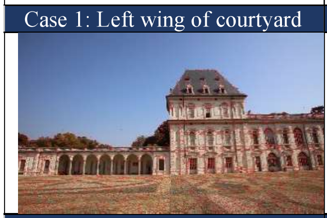
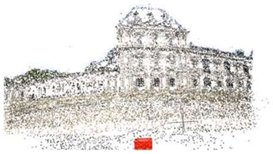
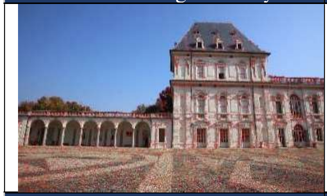
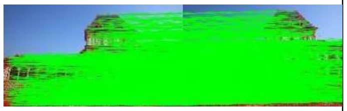
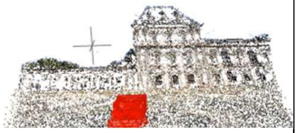
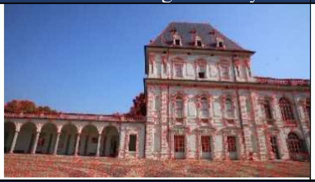
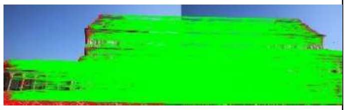
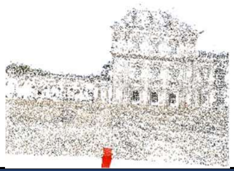
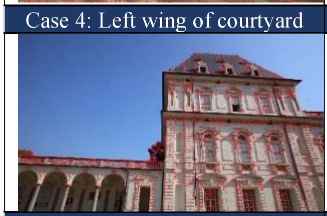
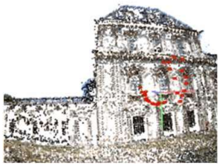
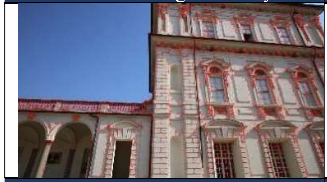
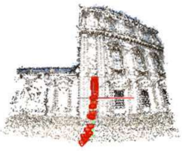
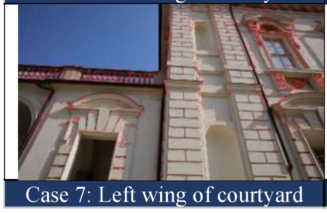
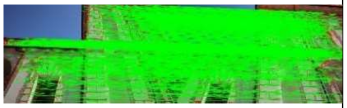
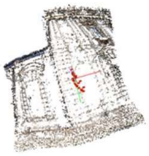
Feature detection and extraction	Feature matching and geometric verification	Structure and motion reconstruction
		
Case 1: Left wing of courtyard	Camera Motion: Tilting	Number of video frames: 49
		
Case 2: Left wing of courtyard	Camera Motion: Tilting	Number of video frames: 34
		
Case 3: Left wing of courtyard	Camera Motion: Tilting	Number of video frames: 28
		
Case 4: Left wing of courtyard	Camera Motion: Tilting	Number of video frames: 32
		
Case 5: Left wing of courtyard	Camera Motion: Tilting	Number of video frames: 33
		
Case 6: Left wing of courtyard	Camera Motion: Tilting	Number of video frames: 34
		
Case 7: Left wing of courtyard	Camera Motion: Tilting	Number of video frames: 32

Figure 17. Results of the three steps of the COLMAP SfM pipeline for the iterative reconstruction. Processing of the frame extracted from the video sequence of the left wing of the courtyard of the Valentino Castle shot with tilting camera motion. The sequence 1 was shot at a distance of 43 m, the sequence 2 at 38 m, the

sequence 3 at 33 m, the sequence 4 at 28 m, the sequence 5 at 18 m, the sequence 6 at 10 m, the sequence 7 at 5 m.

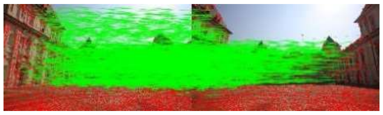
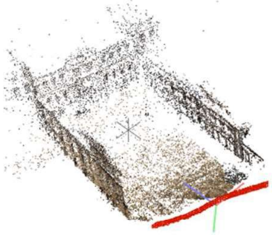
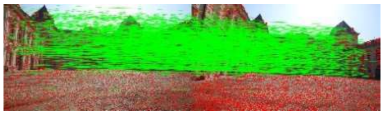
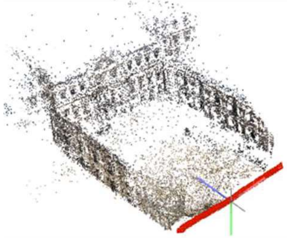
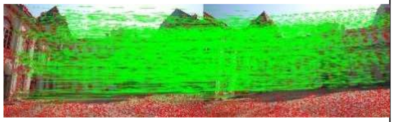
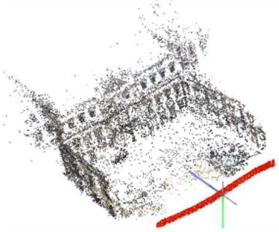
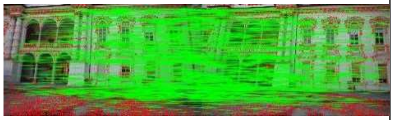
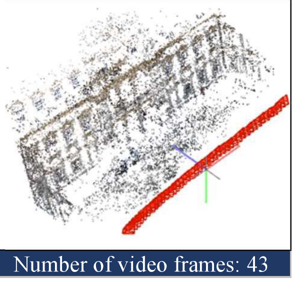
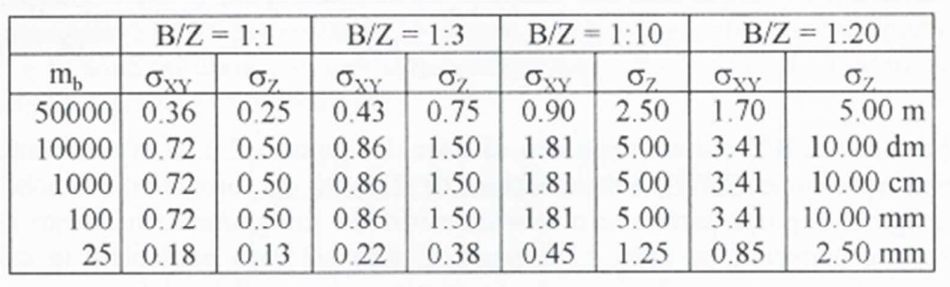
Feature detection and extraction	Feature matching and geometric verification	Structure and motion reconstruction
		
Case 1: Façade	Camera Motion: Trucking	Number of video frames: 41
		
Case 2: Façade	Camera Motion: Trucking	Number of video frames: 42
		
Case 3: Façade	Camera Motion: Trucking	Number of video frames: 35
		
Case 4: Façade	Camera Motion: Trucking	Number of video frames: 43

Figure 18. Results of the three steps of the COLMAP SfM pipeline for the iterative reconstruction. Processing of the frame extracted from the video sequence of the façade of the Valentino Castle shot with trucking camera motion. The sequence 1 was shot at a distance of 85 m, the sequence 2 at 65 m, the sequence 3 at 45 m and the sequence 4 at 25 m.

2.3.4 Metric quality assessment and evaluation results

The analysis examined the results obtained from the previous processing step in order to assess the precision and the accuracy of the models and to evaluate their metric quality.

Table 2. Correlation of the precision in terms of Standard Deviation (σ) related to the frame scale factor (m) and the value of the base ratio (B/Z) (Kraus and Waldhäusl, 1990).



	B/Z = 1:1		B/Z = 1:3		B/Z = 1:10		B/Z = 1:20	
m_b	σ_{XY}	σ_Z	σ_{XY}	σ_Z	σ_{XY}	σ_Z	σ_{XY}	σ_Z
50000	0.36	0.25	0.43	0.75	0.90	2.50	1.70	5.00 m
10000	0.72	0.50	0.86	1.50	1.81	5.00	3.41	10.00 dm
1000	0.72	0.50	0.86	1.50	1.81	5.00	3.41	10.00 cm
100	0.72	0.50	0.86	1.50	1.81	5.00	3.41	10.00 mm
25	0.18	0.13	0.22	0.38	0.45	1.25	0.85	2.50 mm

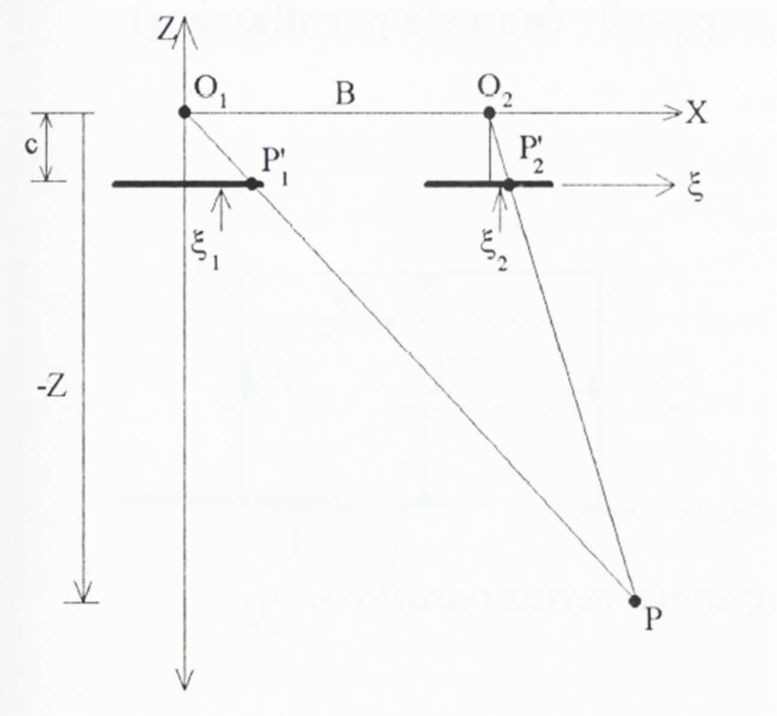


Figure 19. Geometric sketch of two cameras perpendicular to the baseline and parallel each other (Kraus and Waldhäusl, 1990).

In order to perform this evaluation, some photogrammetric rules are reported following.

In photogrammetry, the precision is related to the frame scale factor (m) and the value of the base ratio B/Z (with B for the baseline).

In a model that follows the Gaussian distribution, the precision coincides with Standard Deviation.

The following Table (Kraus and Waldhäusl, 1990) shows this correlation in terms of Standard Deviation (σ), from which derives that:

- For the same base ratio, the standard deviation of X, Y and Z are directly proportional to the frame scale factor;

- For the same frame scale, Residuals in Z are inversely proportional to the base ratio. While Residuals in X and Y slowly increase to decreasing of the base ratio.
- For a specific base, Residuals in Z increase with the square of the distance between camera and object (Kraus and Waldhäusl, 1990).

Consequently, for high overlapping between consecutive frames the precision dramatically decreases.

Precision analysis

In order to analyse the precision of the point cloud obtained processing frame of the videos, the values of the average of the reprojection Residuals overall image observations, expressed in pixel, from the bundle adjustment report of the SfM process were examined (in COLMAP also called Final Cost). All values of Residuals for each case were used for the estimation of the Mean and the Standard Deviation. Moreover, the Minimum and Maximum values of Residuals were highlighted. All these values expressed in pixel were transformed in centimetre with the Ground Sample Distance (GSD) calculation, according to the corresponding taken the distance. The GSD in an image varies because there are parts that are closer to and others that are further away from the camera. In this thesis, the GSD is always evaluated at the point of the object closest to the camera. After the transformation in centimetre, the decimal values are neglected. The results are set out in Table 3 for the tilting case.

Table 3. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding tilting camera motion case.

Camera motion	Mean [px]	Standard Deviation [px]	Min Residual [px]	Max Residual [px]	Distance [m]	GSD [cm/px]	Mean [cm]	Standard Deviation [cm]	Min Residual [cm]	Max Residual [cm]
Tilting1	0.4	0.1	0.1	0.6	43.0	4.0	1.5	0.5	0.5	2.3
Tilting2	0.5	0.2	0.1	0.7	38.0	3.6	1.7	0.6	0.4	2.5
Tilting3	0.5	0.1	0.1	0.7	33.0	3.1	1.4	0.4	0.4	2.0
Tilting4	0.5	0.1	0.1	0.7	28.0	2.6	1.3	0.4	0.3	1.8
Tilting5	0.5	0.1	0.1	0.9	18.0	1.7	0.9	0.2	0.2	1.5
Tilting6	0.5	0.1	0.1	0.7	10.0	0.9	0.5	0.1	0.1	0.7
Tilting7	0.6	0.1	0.2	0.7	5.0	0.5	0.3	0.1	0.1	0.3

The problem with the tilting camera motion is that presents high overlapping between consecutive frames which generates a fall of the precision.

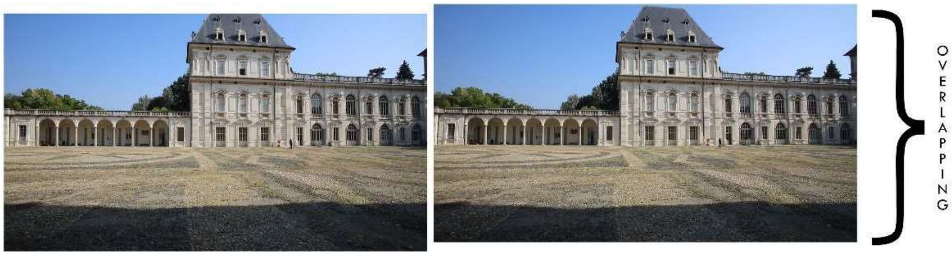


Figure 20. An example of two consecutive frames for the tilting1 case in which the overlapping is over 90% and the precision dramatically decreases.

However, what stands out for the tilting camera motion case in Table 3 is the continual decrease of the values of the Residual. The trend is almost linear and depends on the taken distance. Increasing the distance from which shooting the video from the object, the Residuals increase, even if in a moderate way. This is quite obvious since the camera is more close to the object and the video quality of the pixel is better. However, also if the object is taken from a higher distance, the Residual values are of the order of 2 cm, the result acceptable referring to the architectural case.

The following graphs in Figure 21, 22, 23, 24, 25, 26, 27 show the trend of Residuals values for each distance that follow the Gaussian Distribution.

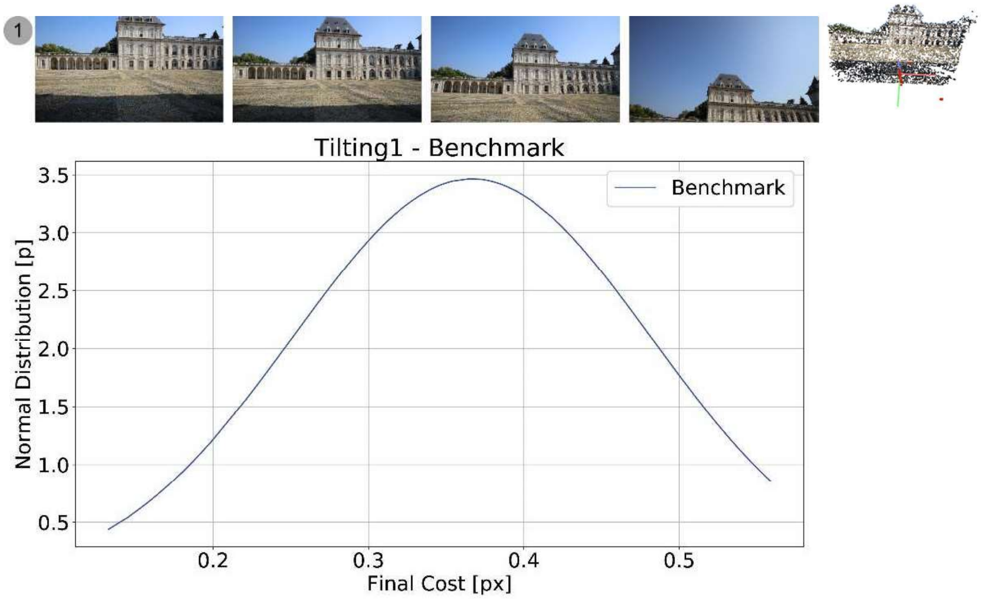


Figure 21. Gaussian distribution of the Residuals values for the tilting1 case, video taken at a distance of 43 m.

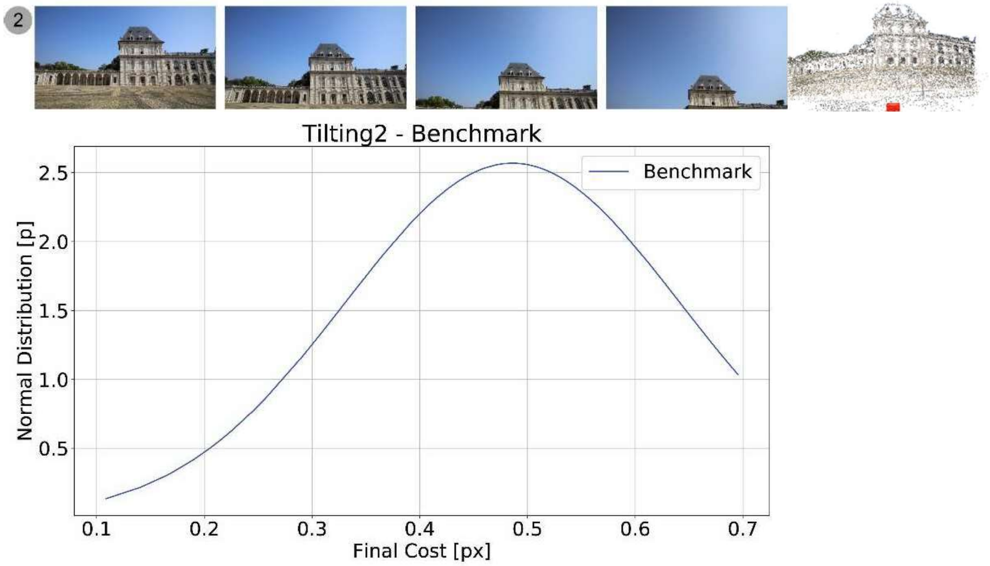


Figure 22. Gaussian distribution of the Residuals values for the tilting2 case, video taken at a distance of 38 m.

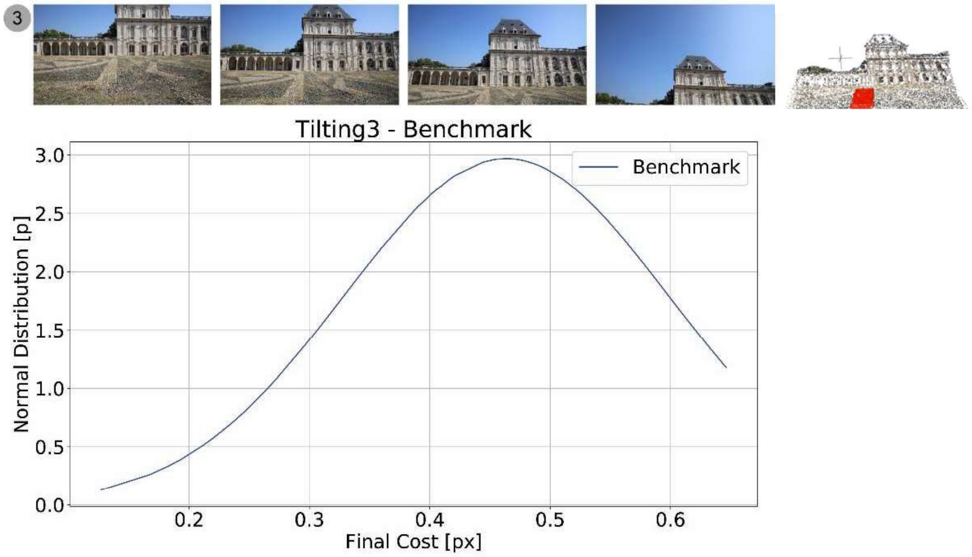


Figure 23. Gaussian distribution of the Residuals values for the tilting3 case, video taken at a distance of 33 m.

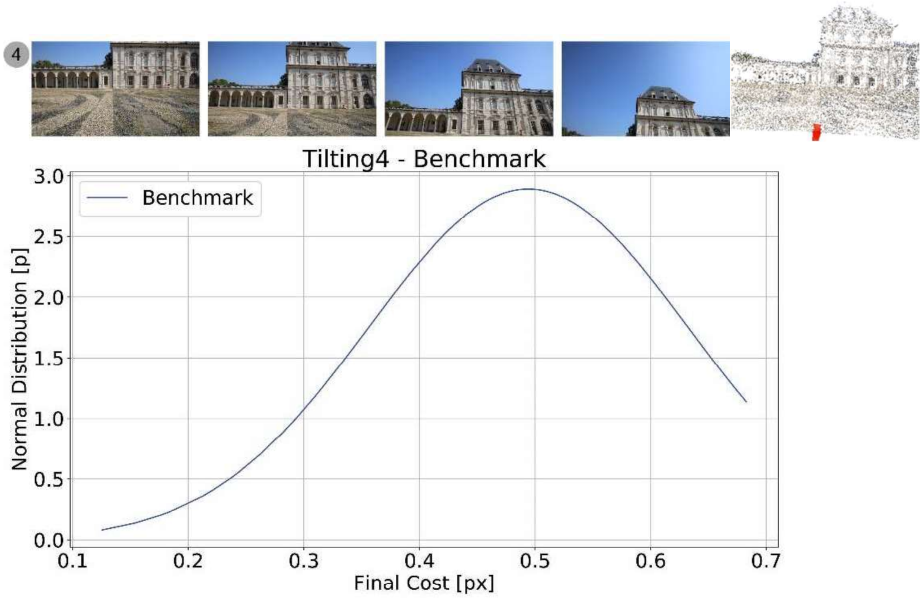


Figure 24. Gaussian distribution of the Residuals values for the tilting4 case, video taken at a distance of 28 m.

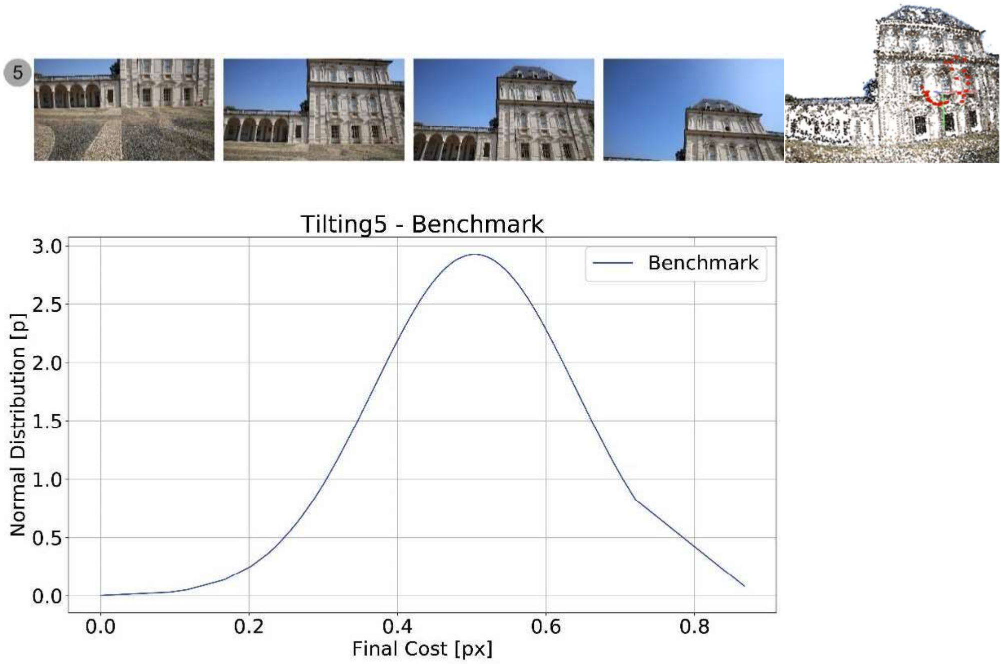


Figure 25. Gaussian distribution of the Residuals values for the tilting5 case, video taken at a distance of 18 m.

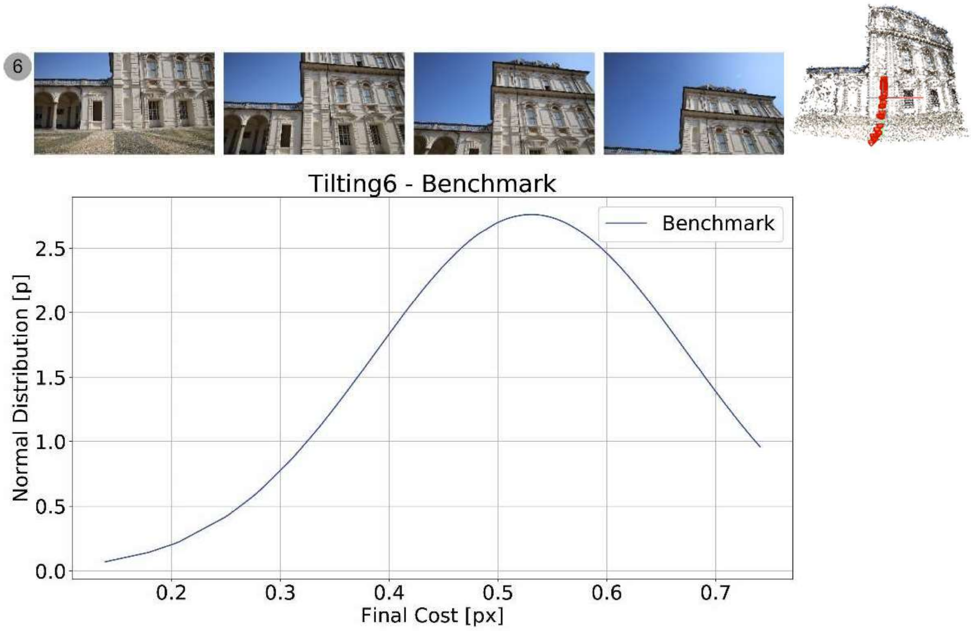


Figure 26. Gaussian distribution of the Residuals values for the tilting6 case, video taken at a distance of 10 m.

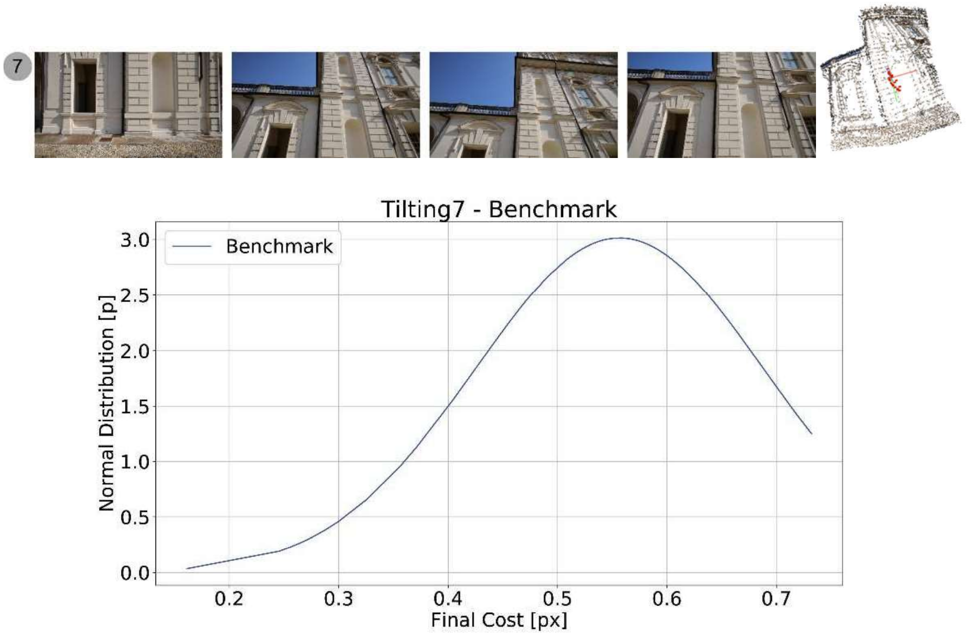


Figure 27. Gaussian distribution of the Residuals values for the tilting7 case, video taken at a distance of 5 m.

Also for the trucking camera motion case, all values of Residuals were used for the estimation of the Mean and the Standard Deviation expressed in pixel and were transformed in centimetre with the Ground Sample Distance (GSD) calculation (considering GSD of the point closest to the camera and rounding the values at the integers). The results are set out in Table 4.

Table 4. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding trucking camera motion case.

Camera motion	Mean	Standard Deviation	Min Residual	Max Residual	Distance	GSD	Mean	Standard Deviation	Min Residual	Max Residual
	[px]	[px]	[px]	[px]	[m]	[cm/px]	[cm]	[cm]	[cm]	[cm]
Trucking1	0.6	0.1	0.3	0.9	85.0	8.0	5.0	0.9	2.3	7.5
Trucking2	0.7	0.1	0.2	0.8	65.0	6.1	3.9	0.5	1.4	4.7
Trucking3	0.7	0.1	0.3	0.8	45.0	4.2	2.9	0.4	1.2	3.4
Trucking4	0.7	0.1	0.2	1.1	25.0	2.3	1.6	0.3	0.6	2.6

Even for the trucking case, there is a linear trend of the values of the Residual that decrease with the reduction of the distance. Moreover, with equal distances, the order of the values of Residuals is the same.

The following graphs in Figure 28, 29, 30, 31 show the trend of Residuals values for each distance that follow the Gaussian Distribution.

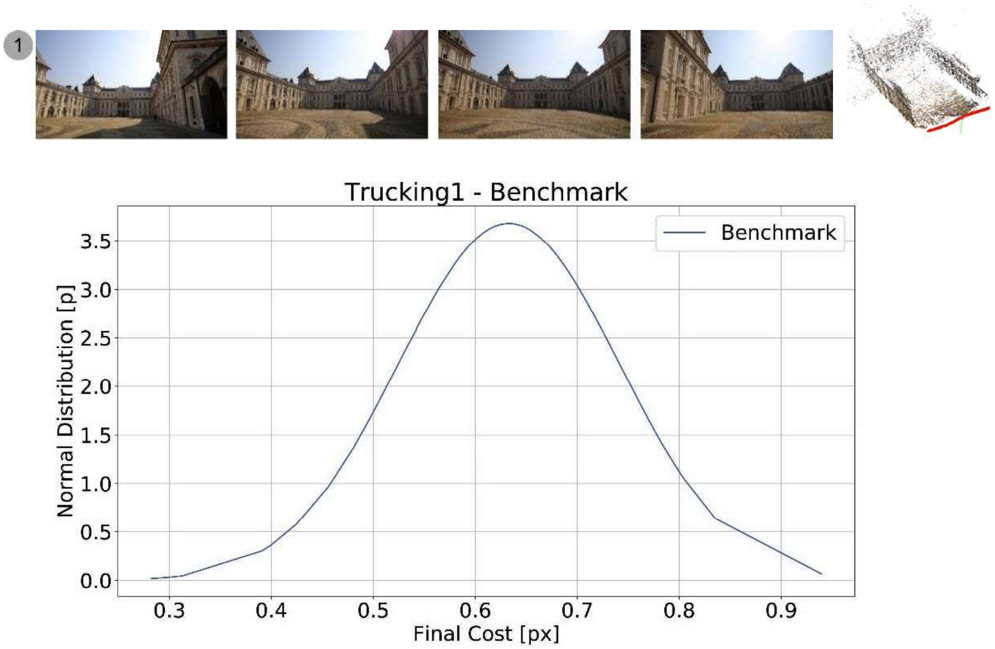


Figure 28. Gaussian distribution of the Residuals values for the trucking1 case, video taken at a distance of 85 m.

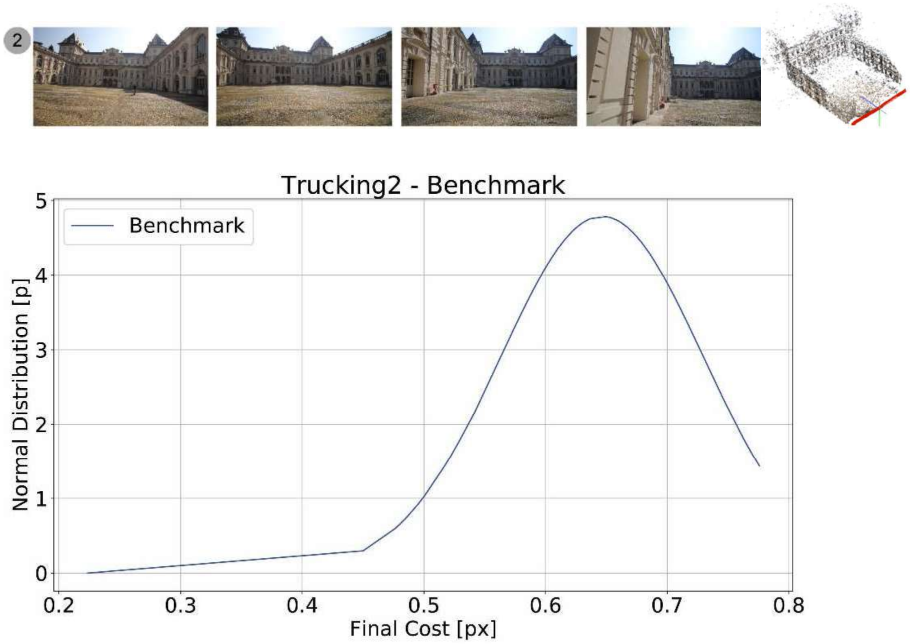


Figure 29. Gaussian distribution of the Residuals values for the trucking2 case, video taken at a distance of 65 m.

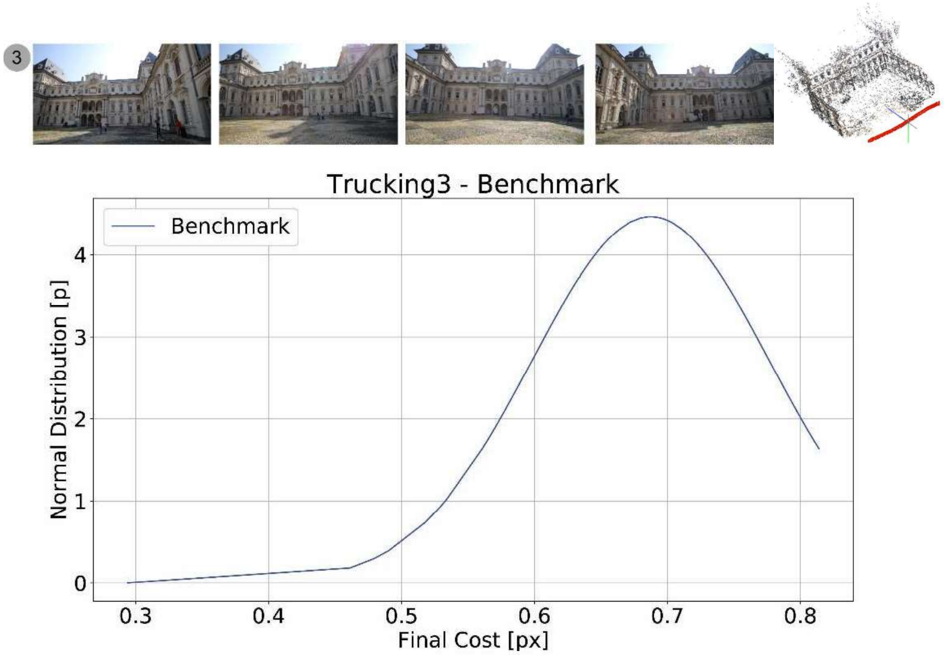


Figure 30. Gaussian distribution of the Residuals values for the trucking3 case, video taken at a distance of 45 m.

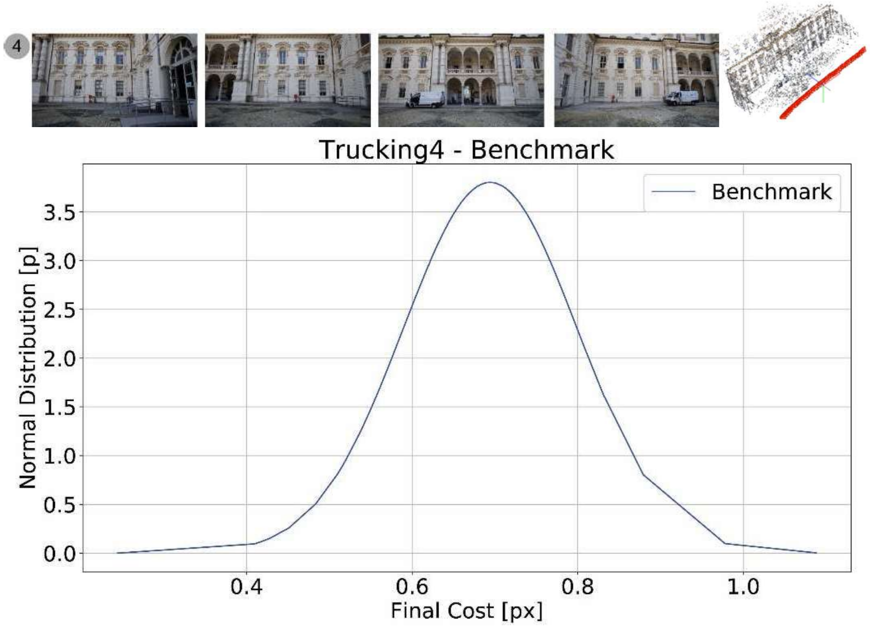


Figure 31. Gaussian distribution of the Residuals values for the trucking4 case, video taken at a distance of 25 m.

Accuracy analysis

To assess the accuracy of the results, the point clouds obtained from the photogrammetric process of the video were compared with a laser scanner survey of the Castle. The point cloud from the laser scanner was chosen as a reference and used to scale the point clouds obtained with the *Alignment and registration* plugin in CloudCompare software.

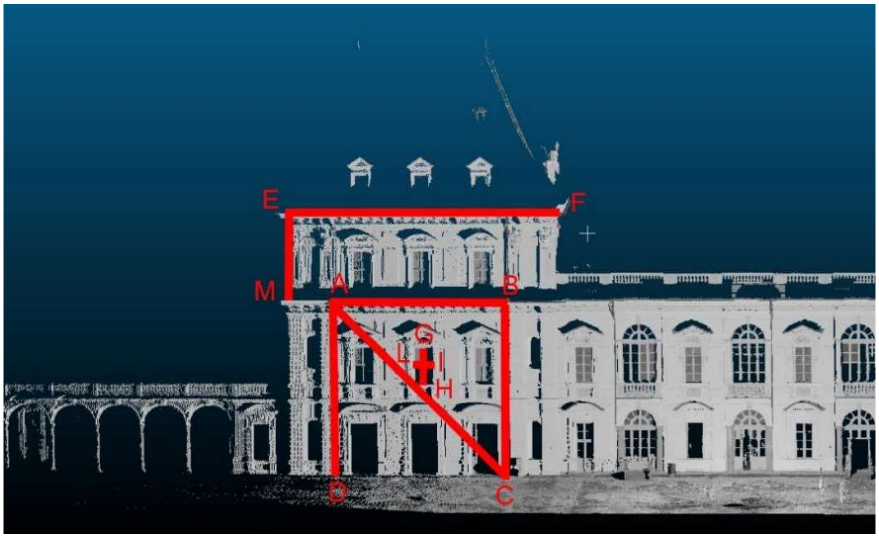


Figure 32. Distances measured on the point cloud obtained from a laser scanner survey of the left wing of the Valentino Castle and used for the comparison with the point cloud of the same wing obtained from the video processing for the tilting camera motion case studies.

To evaluate the accuracy some distances equal in the two point clouds were selected and measured in order to estimate the Residual between the two point cloud. For the tilting case, the distances measured in the two point clouds are shown in Figure 32.

Once measured the same distances in the two point clouds, the estimation of the Residuals between the two measures were reported in Table 5.

As expected, the performance of tilting camera motion from a photogrammetric point of view is affected by the fact that the camera is fixed in a specific point without moving and the frame is acquired with a high overlapping, even if the taking centre moves between frames. The consequence is the presence of high values of Residuals, that get worse decreasing the distance and reach values of even 3 m. This result is also a consequence of the lack of dense points in the laser scanner point cloud chosen as a reference in the comparison.

Table 5. Residuals between the measures of the same distances extracted from the point cloud of the laser scanner, chosen as reference, and the point cloud resulted from the video processing for the tilting camera motion.

		TILTING 1		TILTING 2		TILTING 3		TILTING 4	
Distance	**Laser Scanner**	**COLMAP**	**Residuals**	**COLMAP**	**Residuals**	**COLMAP**	**Residuals**	**COLMAP**	**Residuals**
	[m]	**[m]**	**[m]**	**[m]**	**[m]**	**[m]**	**[m]**	**[m]**	**[m]**
AB	12.51	12.7	-0.19	12.56	-0.05	12.25	0.26	11.84	0.67
AC	17.21	17.93	-0.72	17.89	-0.68	17.15	0.06	20.29	-3.08
AD	11.95	12.07	-0.12	12.11	-0.16	12	-0.05	13.34	-1.39
EF	19.63	19.52	0.11	18.52	1.11	19.21	0.42	16.68	2.95
EM	5.62	5.97	-0.35	5.14	0.48	5.29	0.33	4.32	1.3
FM	20.29	19.93	0.36	19	1.29	19.51	0.78	17.13	3.16
GH	2.49	2.49	0	2.3	0.19	2.63	-0.14	2.58	-0.09
IL	1.26	1.27	-0.01	1.24	0.02	1.3	-0.04	1.16	0.1

After this punctual analysis, a general comparison of the entire point clouds was performed thanks to the Cloud-to-cloud distance comparison (M3C2) plug-in in CloudCompare software (see Section 2.2.3), choosing a maximum local distance of 1 m. The M3C2 algorithm estimated the distance of the video processing point cloud from the reference point cloud considering a maximum distance of 1 m. The scale bar on the right in each comparison shows the colours assigned to the intervals of the distances between 0 and 1 m.

The results are shown in the following Figure 33, 34, 35, 36.

The same limit has emerged in the Cloud-to-Cloud distance comparison, in which the values of distances between the two point clouds increase in the lateral and high part of the point cloud due to the presence of distortions that dramatically influence the results, creating curvatures and deformations.

This reflects expectations, since if the projective rays intersect with a large baseline, the precision increases, while if it is null as in the case of tilting, the lateral parts are less precise and more deformed.

However, in the central part of the façade the values are acceptable for the architectural purpose.

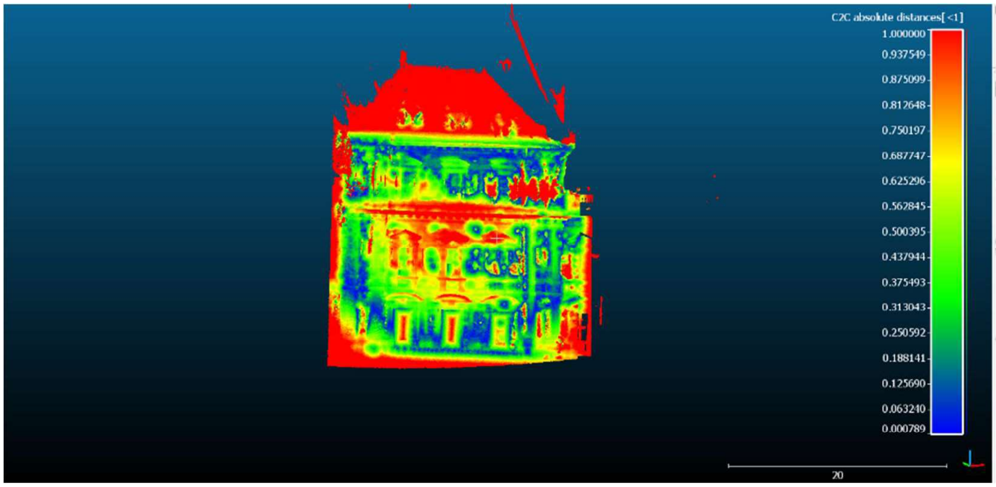


Figure 33. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion tilting1.

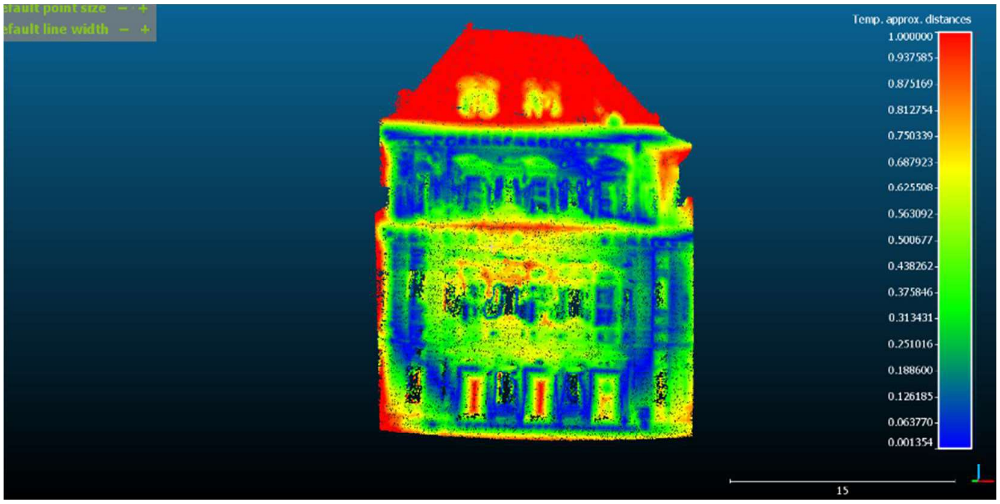


Figure 34. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion tilting2.

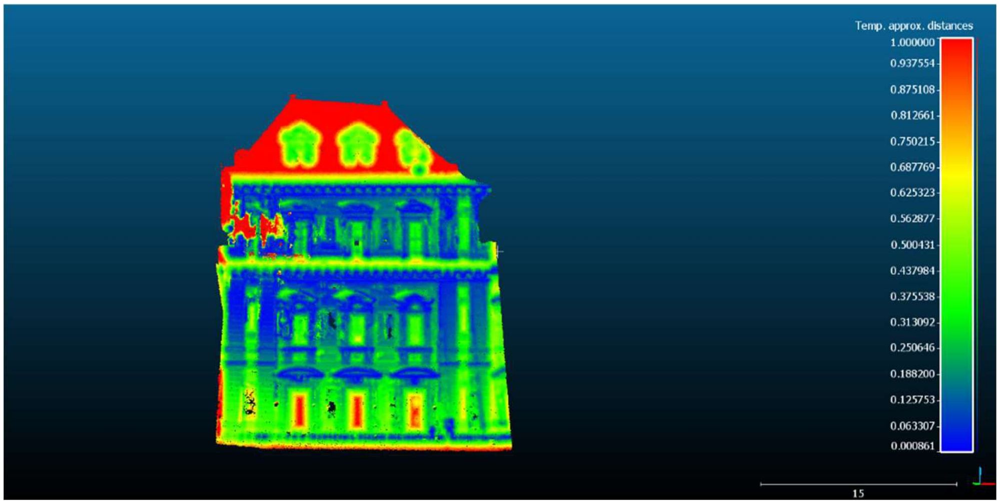


Figure 35. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion tilting3.

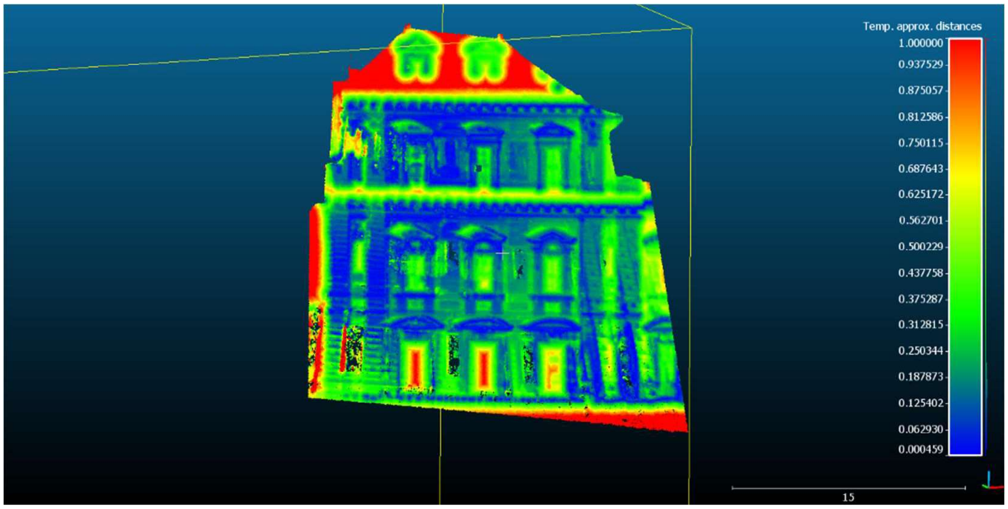


Figure 36. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion tilting4.

Also for the trucking camera motion case, the Residuals between distances (Figure 37) measured on the laser scanner point cloud, chosen as reference, and the point cloud resulted from trucking camera motion analysis, were selected.

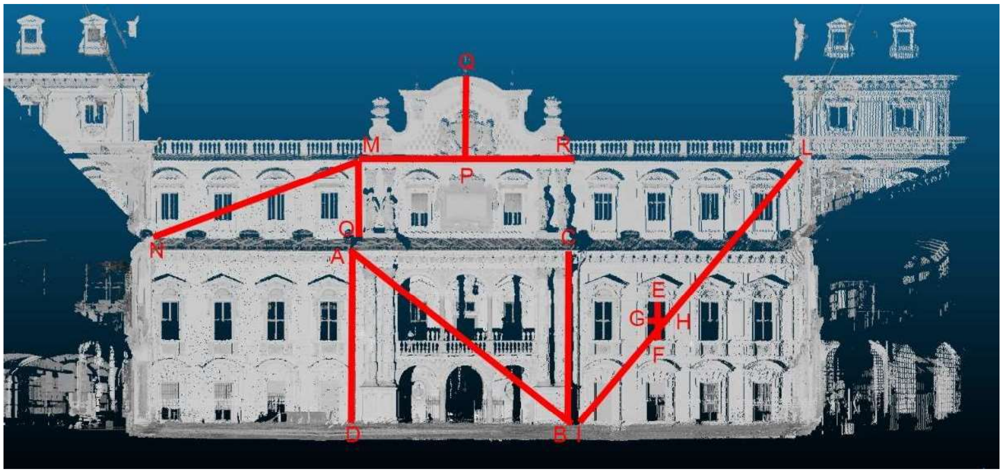


Figure 37. Distances measured on the point cloud obtained from a laser scanner survey of the façade of the Valentino Castle and used for the comparison with the point cloud of the same façade obtained from the video processing for the trucking camera motion case studies.

Table 6. Residuals between the measures of the same distances extracted from the point cloud of the laser scanner, chosen as reference, and the point cloud resulted from the video processing for the trucking camera motion.

		TRUCKING 1		TRUCKING 2		TRUCKING 3		TRUCKING 4	
Distance	Laser Scanner	COLMAP	Residuals	COLMAP	Residuals	COLMAP	Residuals	COLMAP	Residuals
	[m]	[m]	[m]	[m]	[m]	[m]	[m]	[m]	[m]
AB	19.23	19.44	-0.21	19.54	-0.31	19.32	-0.09	19.25	-0.02
AC	15.44	15.54	-0.1	15.21	0.23	15.36	0.08	15.43	0.01
AD	13.02	13.03	-0.01	13.04	-0.02	13.03	-0.01	13.03	-0.01
EF	2.76	2.7	0.06	2.65	0.11	2.73	0.03	2.77	-0.01
GH	1.29	1.29	0	1.28	0.01	1.3	-0.01	1.29	0
IL	23.74	23.91	-0.17	24.32	-0.58	24.07	-0.33		
MN	16.03	16.08	-0.05	16.55	-0.52	16.6	-0.57		
MO	5.39	5.32	0.07	5.55	-0.16	5.15	0.24		
MR	14.67	14.55	0.12	14.66	0.01	14.61	0.06		
PQ	5.31	5.35	-0.04	5.18	0.13	5.06	0.25		

It is clear that moving closer to the façade the point cloud resulted is more accurate and the values of Residual decrease. In each case, the values are not high and are acceptable in the accuracy required by architectural studies.

The comparison of the point cloud from the video processing and the point cloud obtained from the laser scanner, is computed between 0 and 1 m. As shown in the following Figure 38, 39, 40, 41 the distance between the two point clouds is included between 0 and 0.5 m. Moreover, with distance closer to the façade the point cloud results denser.

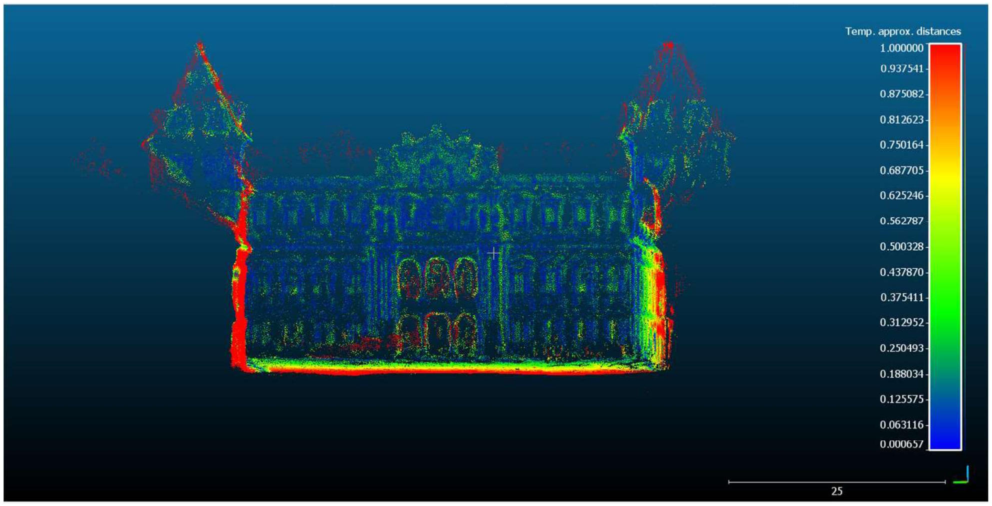


Figure 38. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion trucking1.

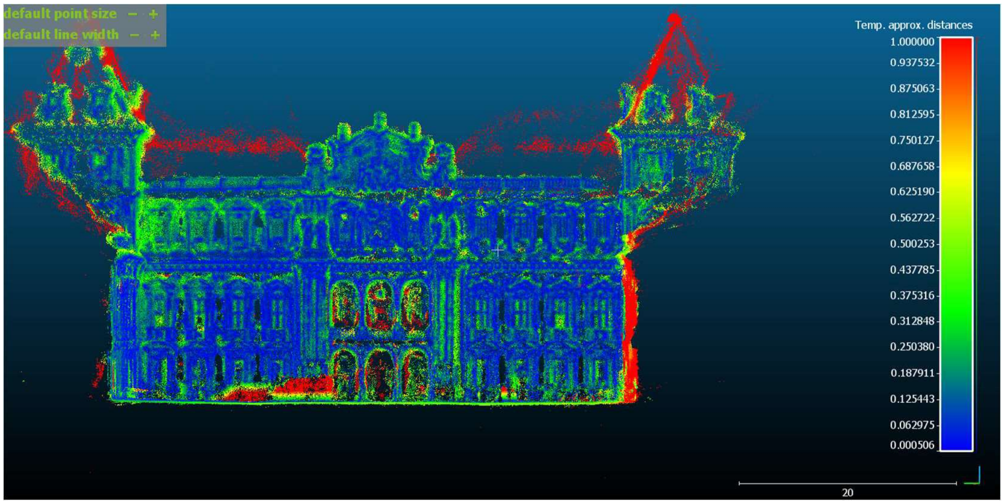


Figure 39. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion trucking2.

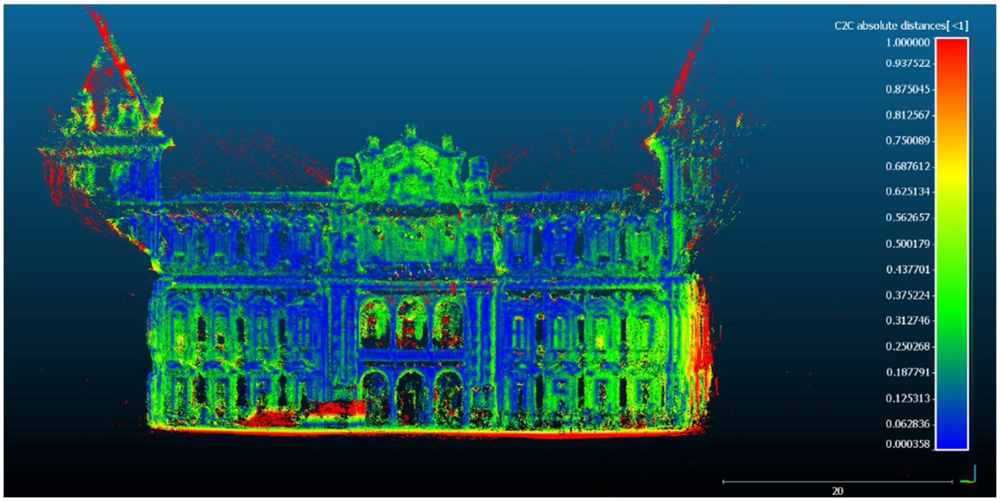


Figure 40. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion trucking3.

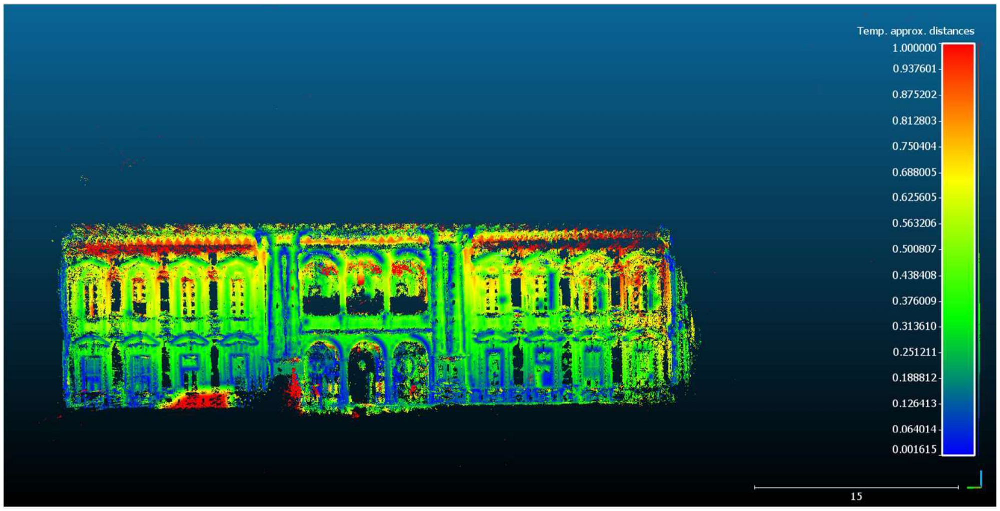


Figure 41. Comparison between laser scanner point cloud, chosen as reference, and the video processing point cloud, for the case of the camera motion trucking4.

The results presented in this section represent the maximum metric quality reachable considering the defined camera motion of the scene taken at specific distances. They constitute the benchmark that will be used in the next parts of the thesis to compare the results of video processing from a metric point of view in the real case of the historical film footage. According to the identified camera motions in the footage and the taking distance of the camera, the precision and accuracy of the results of the photogrammetric reconstruction from the frame can be assessed.

References

Al Khalil, O. and Grussenmeyer, P., 2019. 2D & 3D reconstruction workflows from archive images, case study of damaged monuments in Bosra Al-Sham city (Syria), Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W15, 55–62, https://doi.org/10.5194/isprs-archives-XLII-2-W15-55-2019, 2019.

Bakuła, K., Mills, J. P., and Remondino, F., 2019. A review of benchmarking in photogrammetry and remote sensing, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-1/W2, 1–8, https://doi.org/10.5194/isprs-archives-XLII-1-W2-1-2019, 2019.

Barazzetti, L., Fangi, G., Remondino, F., Scaioni, M., 2010. Automation in multi-image spherical photogrammetry for 3D architectural reconstructions. Proc. of 11th Int. Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2010), Paris, France. doi.org/10.2312/PE/VAST/VAST10S/0 75-081.

Bartos, K., Pukanská, K., & Sabová, J., 2014. Overview of Available Open-Source Photogrammetric Software, its Use and Analysis. International journal for innovation education and research, 2, 62-70.

Bitelli, G., Girelli, V.A., Marziali, M., Zanutta, A., 2007. Use of historical images for the documentation and the metrical study of cultural heritage by means of digital photogrammetric techniques, in: A. Georgopoulos (Ed.), Proceedings of CIPA, XXI Symposium, Athens, Greece. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI5/C53, 8 pages.

Bracci, F., Drauschke, M., Kühne, S., Márton, Z.-C., 2018. Challenges in fusion of heterogeneous point clouds. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci, Vol. XLII-2., pp. 155-162. https://doi.org/10.5194/isprs-archives-XLII-2-155-2018

Bräuer-Burchardt, C. and Voss, K., 2002. Façade Reco.nstruction of Destroyed Buildings Using Historical Photographs. In: Volume XXXIV Part 4, 2002.

Bryson R. Payne, James F. Lay, and Markus A. Hitz. 2014. Automatic 3D object reconstruction from a single image. In Proceedings of the 2014 ACM Southeast Regional Conference (ACM SE ’14). Association for Computing Machinery, New York, NY, USA, Article 31, 1–5. DOI:https://doi.org/10.1145/2638404.2638495.

Byrne, J., O'Keeffe, E., Lennon, D., & Laefer, D.F., 2017. 3D Reconstructions Using Unstabilized Video Footage from an Unmanned Aerial Vehicle. In: J. Imaging, Vol. 3, pp. 15-25.

CloudCompare Version 2.6.1 user manual. 2014. http://www.cloudcompare.org (17 March 2019).

COLMAP, Johannes L. Schoenberger, 2019. COLMAP - Structure-From-Motion and Multi-View Stereo. https://github.com/colmap/colmap (30 June 2019). Commission IV, Symposium 2002 Geospatial Theory, Processing and Applications (July 9-12, 2002, Ottawa, Canada).

Condorelli, F. and Rinaudo, F., 2018. Cultural Heritage reconstruction from historical photographs and videos, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2, 259-265, https://doi.org/10.5194/isprs-archives-XLII-2-259-2018.

Condorelli, F. and Rinaudo, F., 2019. Benchmark of metric quality assessment in photogrammetric reconstruction for historical film footage. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. XLII-2/W11, pp. 443-448, doi.org/10.5194/isprs-archives-XLII-2-W11-443-2019.

Condorelli, F., Higuchi, R., Nasu, S., Rinaudo, F., and Sugawara, H., 2019. Improving performance of feature extraction in SfM algorithms for 3D sparse point cloud, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W17, 101–106, https://doi.org/10.5194/isprs-archives-XLII-2-W17-101-2019, 2019.

Cusicanqui, J., Kerle, N., Nex, F., 2018. Usability of aerial video footage for 3-D scene reconstruction and structural damage assessment. Nat. Hazards Earth Syst. Sci., 18, 1583-1598. doi.org/10.5194/nhess-18-1583-2018.

Dameri, A., 2009. Storia e architettura. Il castello del valentino. Le residenze sabaude, Allemandi Torino, 108-122.

D'Annibale, E., Piermattei, L., Fangi, G., 2011. Spherical photogrammetry as emergency photogrammetry. XXIIIrd International CIPA Symposium, Prague, Czech Republic. ISBN: 978-80-01-04856.

Delis, P., Zacharek, M., Wierzbicki, D., and Grochala, A., 2017. Point Cloud derived from video frames: accuracy assessment in relation to terrestrial laser scanning and digital camera data. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. XLII-2/W3, pp. 217-223, https://doi.org/10.5194/isprs-archives-XLII-2-W3-217-2017.

Fangi, G., 2007. The multi-image spherical panoramas as a tool for architectural survey. XXI International CIPA Symposium, Athens, Greece.

Farella, E. M., Torresani, A., and Remondino, F., 2019. Sparse point cloud filtering based on covariance features. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. XLII-2/W15, pp. 465–472, https://doi.org/10.5194/isprs-archives-XLII-2-W15-465-2019.

Forssén, E., Ringaby, E., 2010. Rectifying rolling shutter video from hand-held devices. CVPR10, San Francisco, USA IEEE Conference on Computer Vision and Pattern Recognition June 2010.

Gerke, M., Nex, F., Remondino, F., Jacobsen, K., Kremer, J. et al., 2016. Orientation of oblique airborne image sets - Experiences from the ISPRS/Eurosdr benchmark on multi-platform photogrammetry. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 41 (2016), S. 185-191. DOI: http://dx.doi.org/10.5194/isprsarchives-XLI-B1-185-2016.

Girardeau-Montaut, D., Roux, M., Marc., R., Thibault, G., 2005. Change detection on points cloud data acquired with a ground laser scanner. In: Workshop "Laser scanning 2005", Enschede, the Netherlands, September 12-14, 2005.

Gruen, A., 1997. Fundamentals of videogrammetry - A review. Human Movement Science 16, Elsevier Science, pp. 155-187.

Gruen, A.; Remondino, F.; Zhang, L., 2004. Photogrammetric reconstruction of the Great Buddha of Bamiyan, Afghanistan. In: The Photogrammetric Record, Volume 19, Issue 107, pp. 177–199.

Hafeez, J., Seunghyun, J., Soonchul K., Alaric, H., 2017. Image Based 3D Reconstruction of Texture-less Objects for VR Contents. In: International Journal of Advanced Smart Convergence. Vol. 6, pp. 9-17. https://doi.org10.7236/IJASC.2017.6.1.9.

Hanke, K.; Moser, M.; Rampold, R., 2015. Historic photos and TLS data fusion for the 3D reconstruction of a monastery altar ensemble. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5/W7. Proceedings of the 25th International CIPA Symposium (31 August – 04 September 2015, Taipei, Taiwan).

Hedborg, J., 2012. Motion and Structure Estimation From Video.

Hedborg, J., Forssén, E., Felsberg, M., Ringaby, E., 2013. Bundle Adjustment for Rolling Shutter Video. Proceedings of SSBA 2013 IAPR, March 2013.

Hemmleb, M., 1999. Digital rectification of historical images. In: IAPRS, Vol. XXXII, CIPA International Symposium, 3-6 October 1999, Olinda, Brasil.

Herráez, J., Martínez, J., Coll, E., Martín, M.; Rodríguez, J. 2016. 3D modeling by means of videogrammetry and laser scanners for reverse engineering. Measurement 87, Elsevier, pp. 216-227.

Hu, C., Pan, Z., Li, P., 2019. A 3D Point Cloud Filtering Method for Leaves Based on Manifold Distance and Normal Estimation. In: Remote Sens, Vol. 11, pp. 198-118, https://doi.org/10.3390/rs11020198.

Huang, Q., Wang, H., Koltun, V. 2015. Single-view reconstruction via joint analysis of image and shape collections. ACM Trans. Graph. 34, 4, Article 87 (August 2015), 10 pages. DOI:https://doi.org/10.1145/2766890.

Jancosek, M. and Pajdla, T., 2011. Multi-view reconstruction preserving weakly-supported surfaces. In CVPR 2011, pp. 3121-3128.

Javaheri, C., Brites, F., Pereira J., Ascenso, 2017. Subjective and objective quality evaluation of 3D point cloud denoising algorithms, 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Hong Kong, 2017, pp. 1-6, doi: 10.1109/ICMEW.2017.8026263.

Khoshelham, K., Díaz Vilariño, L., Peter, M., Kang, Z., and Acharya, D., The ISPRS benchmark on indoor modelling, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W7, 367–372, https://doi.org/10.5194/isprs-archives-XLII-2-W7-367-2017, 2017.

Kniaz, V. V., Remondino, F., and Knyaz, V. A., 2019. Generative adversarial networks for single photo 3D reconstruction, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W9, 403–408, https://doi.org/10.5194/isprs-archives-XLII-2-W9-403-2019, 2019.

Kraus, K., Waldhäusl, P., 1990. Photogrammetrie, Bonn, Dummler.

Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). ISPRS Journal of Photogrammetry and Remote Sensing, 82, pp. 10-26, https://doi.org/10.1016/j.isprsjprs.2013.04.009.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Computer Vision, Vol. 60, pp. 91110.

Luhmann, T., Robson, S., Kyle, S., & Boehm, J., 2019. Close-Range Photogrammetry and 3D Imaging. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110607253.

Maiwald, F., 2019. Generation of a benchmark dataset using historical photographs for an automated evaluation of different feature matching methods, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 87–94, https://doi.org/10.5194/isprs-archives-XLII-2-W13-87-2019, 2019.

Maiwald, F., Henze, F., Bruschke, J., and Niebling, F., 2019. Geo-information technologies for a multimodal access on historical photographs and maps for research and communication in urban history, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W11, 763–769, https://doi.org/10.5194/isprs-archives-XLII-2-W11-763-2019, 2019.

Maiwald, F., Schneider, D., Henze, F., Münster, S., and Niebling, F., 2018. Feature matching of historical images based on geometry of quadrilaterals, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2, 643–650, https://doi.org/10.5194/isprs-archives-XLII-2-643-2018, 2018.

Maiwald, F., Vietze, T., Schneider, D., Henze, F., Münster, S., and Niebling, F., 2017. Photogrammetric analysis of historical image repositories for virtual reconstruction in the field of digital humanities, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W3, 447–452, https://doi.org/10.5194/isprs-archives-XLII-2-W3-447-2017, 2017.

Moulon, P., Monasse, P., Perrot, R. and Marlet, R., 2016. OpenMVG: Open multiple view geometry. In International Workshop on Reproducible Research in Pattern Recognition, pp. 60-74.

Nex, F., Gerke, M., Remondino, F., Przybilla, H.-J., Bäumker, M., and Zurhorst, A., 2015. ISPRS benchmark for multi-platform photogrammetry, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., II-3/W4, 135–142, https://doi.org/10.5194/isprsannals-II-3-W4-135-2015, 2015.

Özdemir, E., Toschi, I, Remondino, F., 2019. A Multi-Purpose Benchmark for Photogrammetric Urban 3D Reconstruction in a Controlled Environment, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

Pavelka, K., Šedina, J., Raeva, P., and Hůlková, M., 2017. Modern processing capabilities of analog data from documentation of the great Omayyad Mosque in Aleppo, Syria, damaged in civil war. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W5, 561-565 https://doi.org/10.5194/isprs-archives-XLII-2-W5-561-2017.

Pavoni, G., Dellepiane, M., Callieri, M., Scopigno, R., 2016. Automatic selection of video frames for path regularization and 3D Reconstruction. EUROGRAPHICS Workshop on Graphics and Cultural Heritage. doi.org/10.2312/gch.20161376.

Pierrot-Deseilligny, M. and Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach: an application to surface reconstruction from SPOT5-HRS stereo imagery. Int. Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, Vol. 36(1/W41).

Poloprutskýa, Z., Fraštiab, M., Marčiš, M, 2019. 3D digital reconstruction based on archived terrestrial photographs from metric cameras. In: Acta Polytechnica 59(4):384–398, 2019, https://doi.org/10.14311/AP.2019.59.0384.

Rahaman, H.; Champion, E., 2019. To 3D or Not 3D: Choosing a Photogrammetry Workflow for Cultural Heritage Groups. Heritage 2019, 2, 1835-1851.

Remondino, F., 2003. Recovering metric information from old monocular video sequences. In: Proceedings of 6th Conference on Optical 3D Measurement Techniques, Gruen/Kahmen editors, pp.214-222 (September 23-25, 2003, Zurich, Switzerland), https://doi.org/10.3929/ethz-a-004665371.

Remondino, F., 2004. Character reconstruction and animation from monocular sequence of images. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV-B5,702-707. XXth ISPRS Congress, Istanbul, Turkey.

Remondino, F., Börlin, N., 2004. Photogrammetric calibration of sequences acquired with a rotating camera. ISPRS Archives, Volume XXXIV-5/W16.

Rodríguez Miranda, Á., Valle Melón, J.M., 2017. Recovering Old Stereoscopic Negatives and Producing Digital 3d Models of Former Appearances of Historic Buildings. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W3, pp. 601-608.

Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C. et al., 2012. The ISPRS benchmark on urban object classification and 3d building reconstruction. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3 (2012), Nr. 1, S. 293-298. DOI: https://doi.org/10.5194/isprsannals-i-3-293-2012.

Rupnik, E., Daakir, M. and Pierrot-Deseilligny, M., 2017. MicMac – a free, open-source solution for photogrammetry. Open geospatial data, software and standards. Vol. 2(14).

Saxena, A., Chung, S.H., Ng, A.Y., 2008. 3-D Depth Reconstruction from a Single Still Image. Int J Comput Vis 76, 53–69 (2008). https://doi.org/10.1007/s11263-007-0071-y.

Schindler, G., Dellaert, F., 2012. 4D Cities: Analyzing, Visualizing, and Interacting with Historical Urban Photo Collections. Journal of Multimedia, Vol. 7, No. 2.

Schönberger J.L., Zheng, E., Frahm, JM., Pollefeys, M., 2016. Pixelwise View Selection for Unstructured Multi-View Stereo. ECCV 2016, Lecture Notes in Computer Science, vol 9907, Springer, Cham.

Schönberger, J. L. and Frahm, J. M., 2016. Structure-from-motion revisited. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Vol. 2016, pp. 4104-4113.

Schönberger, J. L., Frahm, J. M., 2016. Structure-from-motion revisited. IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Vol. 2016, 4104-4113, IEEE Computer Society.

Schöps, T., Schönberger, J.L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M. and Geiger, A., 2017. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In Proc. CVPR, pp. 3260-3269.

Settergren, R., 2020. Resection and Monte Carlo covariance from vanishing points for images of unknown origin, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B2-2020, 487–494, https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-487-2020, 2020.

Snavely, N., Seitz, S.M., Szeliski, R., 2007. Modeling the World from Internet Photo Collections. International Journal of Computer Vision November 2008, Volume 80, Issue 2, pp 189–210, doi 10.1007/s11263-007-0107-3.

Stathopoulou, E.-K., Welponer, M., and Remondino, F., 2019. Open-source image-based 3D reconstruction pipelines: review, comparison and evaluation, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W17, 331–338, https://doi.org/10.5194/isprs-archives-XLII-2-W17-331-2019, 2019.

Sung, B.-Y., and Lin, C.-H., 2017. A fast 3D scene reconstructing method using continuous video. EURASIP Journal on Image and Video Processing, doi 10.1186/s13640-017-0168-3.

Tazir, M., Gokhool, T., Checchin, P., Malaterre, L., Trassoudaine, L., 2018. CICP: Cluster Iterative Closest Point for sparse-dense point cloud registration. In: Robotics and Autonomous Systems, Vol. 108, pp 66-86, doi.org/10.1016/j.robot.2018.07.003.

Torresani, A. and Remondino, F., 2019. Videogrammetry vs photogrammetry for heritage 3D reconstruction, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W15, 1157–1162, https://doi.org/10.5194/isprs-archives-XLII-2-W15-1157-2019, 2019.

Van den Heuvel, F., 2001. Object reconstruction from a single architectural image taken with an uncalibrated camera. Photogrammetrie, Fernerkundung, Geoinformation, 2001 (4), pp 247-260.

Van den Heuvel, F.A., 1998. 3D reconstruction from a single image using geometric constraints. In: ISPRS Journal of Photogrammetry & Remote Sensing 53: 354-368.

Vautherin, J., Rutishauser, S., Schneider-Zapp, K., Fai Choi, H., Chovancova, V., Glass, A., Strecha, C. 2016. Photogrammetric accuracy and modelling of rolling shutter cameras. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp.139-146, (2016).

Verhoeven, Geert. 2016. Basics of Photography for Cultural Heritage Imaging. In 3D Recording, Documentation and Management of Cultural Heritage, ed. Efstratios Stylianidis and Fabio Remondino, 127–251. Caithness: Whittles Publishing.

Waldhäusl P, Ogleby C.L., Lerma J.L., Georgopoulos, A., 2013. 3x3 rules for simple photogrammetric documentation of architecture. URL: http://cipa.icomos.org/wpcontent/uploads/2017/02/CIPA__3x3_rules__20131018. pdf

Wang, J., Xu, K., Liu, L., Cao, J., Liu, S., UW-Milwaukee, Z., Gu, X., 2013. Consolidation of low-quality point clouds from outdoor scenes. In: Proceeding SGP '13. Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing, pp. 2017-2016.

Wong, S., Chan, K., 2010. 3D object model reconstruction from image sequence based on photometric consistency in volume space. In: Formal Pattern Analysis & Applications, Vol. 13, pp. 437-450. https://doi.org/10.1007/s10044-009-0173-y.

Wu, J., Wang, Y., Xue, T., Sun, X., Freeman, W. T. and Tenenbaum, J. B., 2017. MarrNet: 3D shape reconstruction via 2.5D sketches. arXiv.org.

Yang, B., Rosa, S., Markham, A., Trigoni N, Wen, H., 2019. Dense 3D Object Reconstruction from a Single Depth View, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 12, pp. 2820-2834, 1 Dec. 2019, doi: 10.1109/TPAMI.2018.2868195.

Yang, H. T., Zhang, H. B., 2017. Automatic 3D reconstruction of a polyhedral object from a single line drawing under perspective projection, Comput. Graph. 65 (2017): 45-59.

Ying Yang, M., Yilmaz A., 2018. Report for Scientific Initiative 2017 ISPRS Benchmark on UAV Semantic Video Segmentation https://www.isprs.org/society/si/SI-2017/ISPRS-SI2017-TC2_WG5_Yang_Report.pdf.

Zawieska, D., Markiewicz, J. S., Kopiasz, J., Tazbir, J., and Tobiasz, A., 2017. 3D modelling of the Lusatian Borough in Biskupin using archival data. In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W3, 665-669, https://doi.org/10.5194/isprs-archives-XLII-2-W3-665-2017.

Zheng, E., Dunn, E., Jojic, V. and Frahm, J.M., 2014. Patchmatch based joint view selection and depthmap estimation. In Proc. CVPR, pp. 1510- 1517.

Chapter 3

Recovery material suitable for photogrammetry

3.1 A Match-moving method combining AI and SfM algorithms in historical film footage

As introduced, this section[1] focuses on the examination of the use of historical film footage as material for photogrammetric reconstruction of lost or changed heritage. An important aspect of the documentation process, in fact, is the collection of data and information about heritage. Creating new tools for the final user of these data is an appealing research topic especially in the AI domain. In fact, the volume, the size and the variety of historical data lead to some critical factors. The most important is concerned with the manpower needed to organize and search the documents. To solve this problem the application of Deep Learning (DL) gives opportunities to enhance historical archives and retrieval of heritage information.

For this reason, recent research in this area has seen the rapid development of technologies to support the management and analysis of historical data regarding heritage. Deep learning can be used to automate tasks such as processing these large amounts of data and reducing human effort, thus making them more efficient.

When Artificial Intelligence (AI) is combined with techniques that are widely used in the field of Cultural Heritage such as photogrammetry, the documentation process

[1] Part of the work described in this Chapter has also been previously published in Condorelli et al., 2019 and 2020.

can actually be improved. This is shown in this section in which Deep Learning is used to search for suitable material for photogrammetry.

In the workflow proposed in this work a combination of Deep Learning techniques with photogrammetry is presented. DL is used for the retrieval of primary data used as input material in the standard Structure-from-Motion (SfM) pipeline. This workflow intends to find out how to improve ways to search for architectural heritage within a large quantity of unorganized and low-quality video material and to reduce the effort of the operator in the archive in terms of efficiency and time. In order to achieve this purpose, the automatic detection of a specific monument in film footage and its three-dimensional virtual reconstruction is performed using Deep Learning.

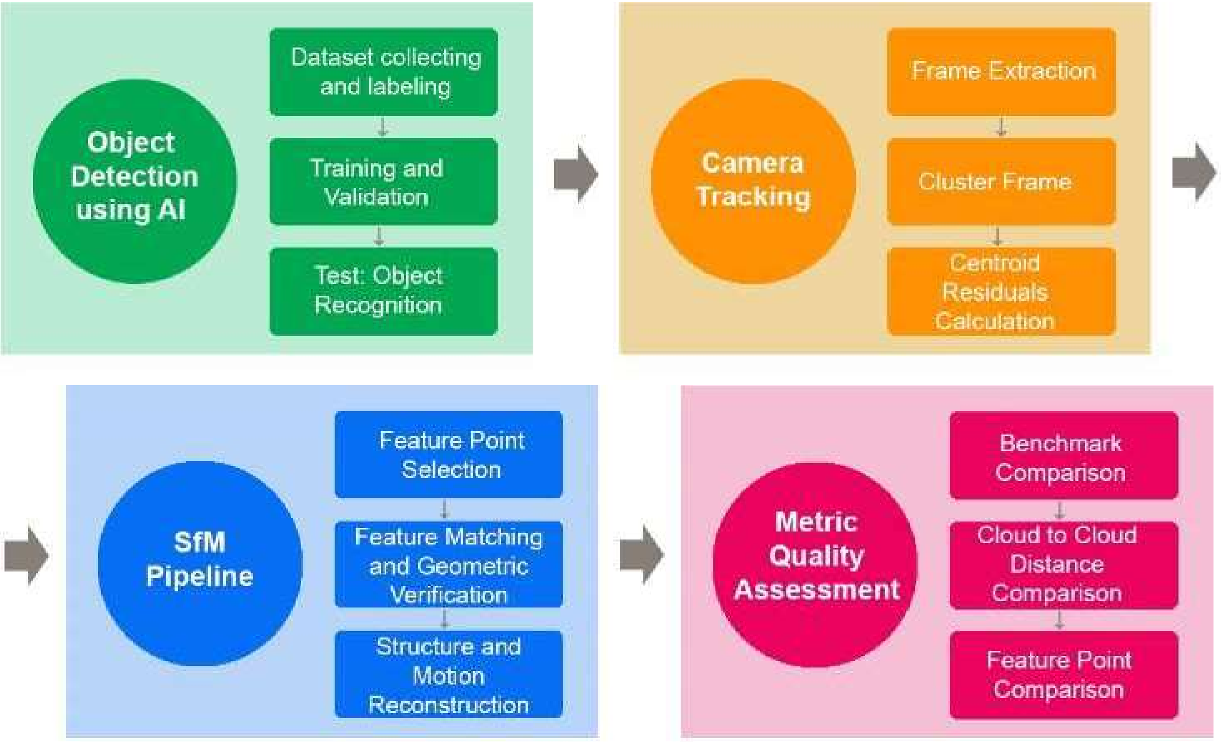

Object Detection using AI
Dataset collecting and labeling
Training and Validation
Test: Object Recognition
Camera Tracking
Frame Extraction
Cluster Frame
Centroid Residuals Calculation
SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction
Metric Quality Assessment
Benchmark Comparison
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 1. The general workflow proposed in this dissertation. The following sections will deal with the application of this workflow on historical film footage, focusing on the two steps of "object detection using AI" and "camera tracking".

In particular the first part of this section deals with the issue regarding the availability and accessibility of these materials in archives, often made difficult by the lack of an appropriate organization of these data. The human effort to find the data of interest represents a significant percentage of the final user work since the indexing of metadata for historical archival material is often incomplete or inaccurate, and the corresponding search engines are therefore not very efficient. The need to identify the object of interest within the amount of material that potentially contains it without the effort of manually examining individual videos in archives has led to the development of an algorithm for automatic monument identification in film footage using Neural Networks.

While the second part tackles the problem concerning the camera movements used to shoot the videos, since historical film footage was not taken for use in 3D reconstruction. To solve this problem an algorithm of automatic identification of camera motion is presented.

Referring to the workflow presented in Chapter 1 (Figure 1), implementing it on historical film footage turns into an innovative match-moving method. Match-moving is a technique used to track the movements of a camera in a 3D space using the images that it acquires while moving. This method is widely used in computer vision, the film industry and video editing, as it makes it possible to match the real scene with virtual creations such as visual effects. Structure from Motion (SfM) is the main part of this process which allows the extraction of the 3D information from the scene. In the workflow here proposed, match-moving allows the exploitation of Artificial Intelligence and SfM algorithms to identify the frames extracted from film footage in which the lost monument appears and that are suitable to be processed with photogrammetry for its 3D reconstruction.

3.1.1 Existing match-moving methods

Camera tracking technology is based on the SfM method, since the determination of the camera position and the field of view is done by analyzing the film shot and extrapolating the 3D data from the original 2D imagery (Condell and Moore, 2006; Zhang et al., 2009; Ingwer et al., 2015).

According to previous studies (Lee et al., 2006), camera match-moving approaches can be divided into two categories: feature-based approaches and model-based approaches. The first uses appropriate feature points, i.e. a method based on a 3D plane tracking technique that allows the estimation of the homographies induced by a 3D plane between successive image pairs (Lourakis and Argyros, 2005); a practical real-time camera tracking system that provides an offline process for space abstraction using features and an online step for feature matching (Dong et al., 2009); and a non-consecutive feature tracking framework for matching interrupted tracks distributed in different subsequences or even in different videos (Zhang et al., 2016). The second uses a known geometric object in a given environment, i.e., the development of a marker-based real-time feature tracking method that operates in unknown environments and uses a known marker for fast recognition and tracking of feature points (Lee et al., 2006). Another technique uses depth information to evaluate camera position and trajectory (Luo et al., 2016).

Various tools and software are available to replicate the camera trajectory, which differ in price, usage, functionality and user interface. Among the commercial software, the most used are Boujou (Vicon Motion Systems Ltd. UK), which uses frame-by-frame comparison to track the camera; and SynthEyes (Andersson Technologies LLC), which can determine how the real camera moved during the shooting, what the field of view (focal length) of the camera was and where the different positions were in 3D space. An example of open source software is ACTS (Zhang et al., 2009), an automatic camera tracking system that can track camera movement but is limited to only two types, pure rotation and free movement. Another limitation is that it only works with long sequences.

Research has been conducted on match-moving for building analysis (Dağlar et al., 2011) and on the use of video material as a source of metric data for filmed architecture (Mancini et al., 2013). Previous research has proposed an automatic workflow for 3D reconstruction of objects of interest in videos that have been captured by simple users mainly for personal or touristic purposes and that for this reason contain noise or information not related to the object to reconstruct. The frames with the object were extracted using a video summarization algorithm and modelled with SfM algorithms (Doulamis, 2018).

However, in the field of architecture a precise pipeline for the virtual reconstruction of lost Cultural Heritage has not yet been defined.

Two important questions arise from the studies reviewed so far. Most of them are outdated and there is a lack of experiments with the most effective Artificial Intelligence techniques, which are very useful for such tasks. In addition, there is no open source software, except in one case that has certain limitations, including the inability to handle short tracks and the tilting and trucking camera motion types, both of which are very common in historical film footage.

3.1.2 Innovation of the proposed workflow

Overall the match-moving process is similar in every software and consists in the following steps, as shown in Figure 2: feature identification and tracking, camera tracking and 3D modelling (Haji et al., 2016; Dağlar et al., 2011).

- Feature tracking consists in determining the position of points of interest in the footage by calculating their motion vectors frame by frame.
- Camera tracking finds the motion of the camera in 3D space by extracting its characteristics (orientation, position and focal length) using SfM.
- 3D modelling is performed with the use of SfM to reconstruct a 3D scene.

However, the standard method has been significantly modified to improve it for a more efficient use for the heritage. In this work, in fact, the goal to be achieved is different from the original match-moving process, which aims to correctly insert an object into a video. This research aims to extract images from historical videos in a way suitable for the photogrammetric process. For this purpose it is necessary to know how the camera was moved to film the video, as this strongly influences the results of the photogrammetric reconstruction. For this purpose the following innovations have been introduced into the proposed workflow:

1) the use of Artificial Intelligence object recognition algorithms as a method for feature tracking, which, as the state of the art of the study shows, is a further development compared to previous studies;
2) the algorithm for camera tracking,
3) the open source SfM algorithms and the metric quality assessment, which certifies the quality of the 3D reconstruction.

This last step was deeply treated in the previous Chapter 2. Following the object detection and the camera motions identification steps will be explained in more detail.

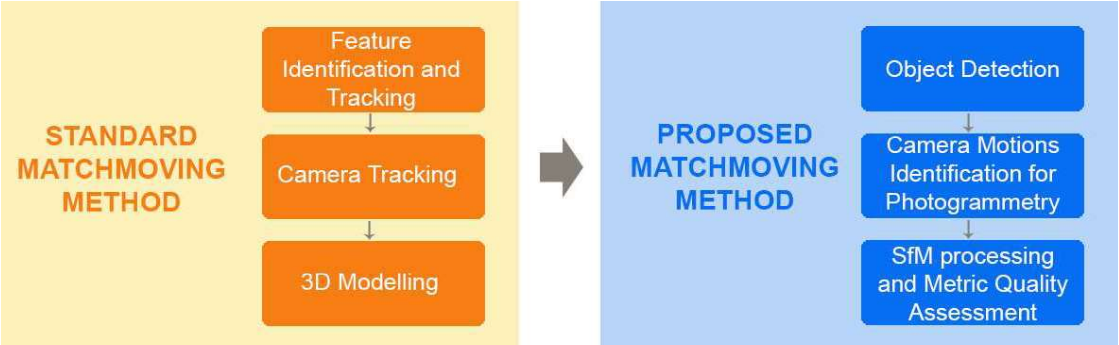

STANDARD MATCHMOVING METHOD
Feature Identification and Tracking
Camera Tracking
3D Modelling
PROPOSED MATCHMOVING METHOD
Object Detection
Camera Motions Identification for Photogrammetry
SfM processing and Metric Quality Assessment

Figure 2. Workflow of the standard match-moving method that has been modified to improve it for a more efficient use for the heritage. A new workflow is proposed introducing the use of AI, the camera tracking algorithm, the use of open source SfM algorithms and the metric quality assessment of the results.

3.2 Architectural heritage detection using Neural Networks

The first step of the workflow is to identify and to track features from the video. This was performed using an object detection Neural Network trained to automatically recognise the monument in the film footage.

Object detection involves the recognition of the searched object by segmenting a region of interest, classifying it by putting a bounding box around it and assigning it a label with the name of the corresponding class. In order to track the object in the video sequence, the evolution of the position of the bounding boxes over time is analysed in order to precisely locate the object.

Object detection is a good solution for applications such as the recognition of monuments in film footage because it allows the tracking of the object even if the image is noisy, the camera is not stable and the object has a complex structure (Parekh et al., 2014).

3.2.1 State of the Art: Artificial Intelligence for Cultural Heritage

Machine Learning technique has become fundamental not only in the field of computer science research, but also in everyday life, finding applications for example in web search engines, fraud detection systems, spam filters, automatic text analysis systems, and medical diagnostic systems. One of the reasons for this growing importance is the successful application of DL methods in areas such as image classification (Krizhevsky et al. 2015, Szegedy et al., 2015; Simonyan and Zisserman, 2015), in which convolutional neural networks (CNNs) exceed the human level in object recognition and image search (Radenović et al., 2016; Tolias et al., 2016).

Besides the improvements of Machine Learning techniques, hardware development, in particular the use of Graphical Processing Units (GPUs), has given a boost to the computational efficiency of such algorithms.

However, only a few studies in the architectural and Cultural Heritage field have developed. Different algorithms within DL, Supervised, Semi-supervised and Unsupervised have been extensively used in Cultural Heritage applications (Fiorucci et al., 2020). So far, thanks to this approach, researchers have been able to classify interesting objects in images of buildings of architectural value (Llamas et al., 2016); identify different monuments based on the feature of the images of monuments (Saini et al., 2017); automatically annotate the cultural assets based on their visual feature and the available metadata (Belhi et al., 2018); recognize a character in images of artworks and their contexts (Montoya Obeso et al., 2019); interpret deep features learned by Convolutional Neural Networks for city recognition (Shi et al., 2019); develop a mobile app to perform monument recognition using convolutional neural networks and to query a database and to extract all the information related to that object (Andrianaivo et al., 2019).

Referring to an urban context, some researchers have concentrated on visual place recognition to help humanity and architecture studies to retrieve information about cities and locations. Training a NN it is, in fact, possible to find out where a picture of a part of a city was taken and to recognise the place querying a database, or predict a geo-location from an image (Khademi et al., 2018). An example is the training of a model that classify images from Tokyo and Pittsburgh and generate the visual explanations and descriptors for each image (Shi et al., 2019) and the investigation of the interpretation of deep features learned by convolutional neural networks for city recognition (Shi et al., 2019). Another study found visual elements (both architectural and not) of a place, such as the city of Paris. These local geo-informative features are queried from a large database of photographs from a particular place offered by Google Street View (Doersch et al., 2012).

In addition to these studies, existing research recognizes the important role played by historical data in archives and the potentialities of DL and proposed different methods: to automatically index and label the documents and search through the collections (Picard et al., 2015); to retrieve images and information on heritage (Yasser et al., 2017) and iconographic contents representing landscapes of the French territory (Gominski et al., 2019). However, the majority of these works consider the analysis of paintings, drawings, images, while film footage has been hardly explored with these techniques. Only an example of Deep Learning application to extract semantic features to analyse the role of intertitles in early cinema was conducted (Bhargav et al., 2019).

Considering these previous studies, the lack of an algorithm to recognise monument in film footage is highlighted. In the following section the proposed solution is presented in detail.

3.2.2 Object Detection Using Neural Networks

The first part of the workflow concerns the use of Neural Networks to recognise the object of interest in the film footage. Among the different types available, Neural Networks that support the object detection algorithm were selected. This solution proved to be effective for the experimented pipeline since it allows image classification even in complex images and with the extraction of bounding boxes of the object recognized.

The usability of the workflow by the operator in the archive is an important aspect to be considered. For this reason, the Luminoth software (https://github.com/tryolabs/luminoth, accessed on July 2020) based on TensorFlow (http://tensorflow.org, accessed on July 2020) was chosen because it implements an object detection algorithm through state-of-the-art networks. In particular, in this work the networks following described are used.

Faster R-CNN Neural Network

Faster R-CNN (Ren et al., 2016), stands for Faster Region-based Convolutional Neural Network, is a deep convolutional network used for object detection, that appears to the user as a single, end-to-end, unified network. The network can accurately and quickly predict the locations of different objects. It is the faster evolution of an R-CNN network (Girshick et al., 2014) the aim of which is to reduce the problem of object detection to a classification problem, which is performed on limited regions of an image. The idea behind this type of network is very simple: sub-portions of an image (regions) are selected and these regions are used as the input of a classifier that uses convolution networks to determine the class of the extracted object. From the computational point of view, it would not be possible to apply the classifier to every possible sub-image of the starting image; for this reason R-CNN was designed to reduce the number of possible regions to be used by the classifier. The R-CNN network uses an algorithm for selecting possible regions (region-proposal) which reduces, around 2000 times, the number of images fed to the classifier. For each proposed region, the classifier that determines the class of the region is applied; and possibly a regression over a set of bounding boxes is applied to determine the optimal bounding box of the region containing the object. For the selection of the region module, a variety of methods for generating category-independent region proposals exist. The main aspect of the Faster R-CNN network is the replacement of the region selection algorithm (a computationally expensive part) with a convolutional network called the Region Proposal Network. The result is a network hundreds of times faster than the R-CNN but with a comparable accuracy. A simplified sketch of the Faster R-CNN network is provided in Figure 3, in which the Region Proposal Network (in green) generates region proposals and for all region proposals in the image, a fixed-length feature vector is extracted from each region using the ROI Pooling layer (in blue). The extracted feature vectors are then classified using the Fast R-CNN and the class scores of the detected objects in addition to their bounding-boxes are returned.

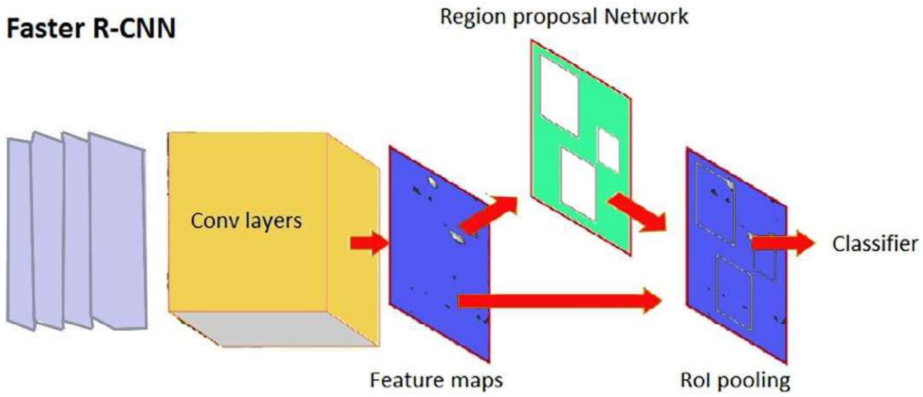


Figure 3. Sketch of the Faster Region-based Convolutional Neural Network (Faster R-CNN).

SSD Neural Network

SSD (Liu et al., 2016) stands for Single Shot Detector and it is oriented to reduced computational demand while keeping an adequate accuracy, is in fact designed for object detection in real-time. The SSD model is simpler if compared to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stage and encapsulates all computation in a single network. To recover the drop in accuracy, SSD applies a few improvements including multi-scale features and default boxes. These improvements allow SSD to match the Faster R-CNN's accuracy using lower resolution images, which further pushes the speed higher. The SSD object detection composes of 2 parts, the extraction of feature maps, and the application of convolution filters to detect objects. The network allows discretizing the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape.

A simplified sketch of the SSD network is provided in Figure 4. As shown in the figure, SSD's architecture builds on the VGG-16 architecture, but discards the fully connected layers. The reason VGG-16 was used as the base network is because of its strong performance in high quality image classification tasks and its popularity for problems where transfer learning helps in improving results. Instead of the original VGG fully connected layers, a set of auxiliary convolutional layers (from conv6 onwards) were added, thus enabling to extract features at multiple scales and progressively decrease the size of the input to each subsequent layer.

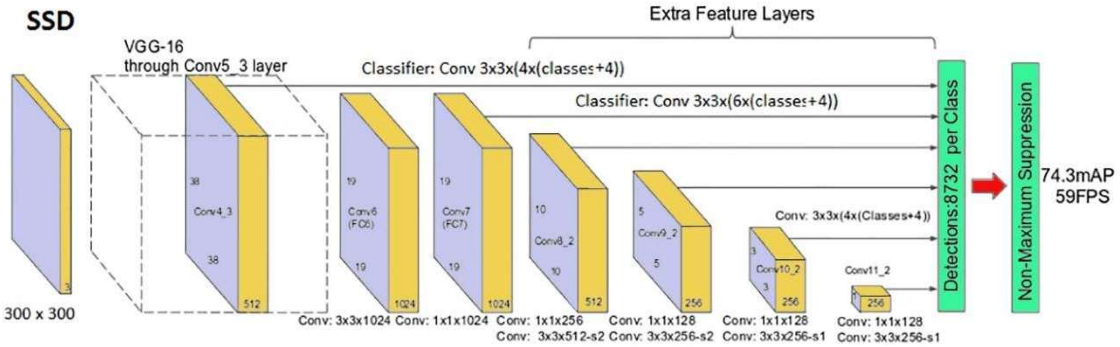


Figure 4. Sketch of the Single Shot Detector (SSD) network.

As a general rule, SSD networks are usually expected to be faster but less accurate than Faster R-CNN networks. This behaviour, however, depends on the sizes of the considered objects and other factors, and it will be discussed in Chapter 4 in the context of investigated Cultural Heritage cases.

The described networks are provided already pre-trained by Luminoth. However, it is possible to add a new element to detect with a further training phase. This point is particularly important for the Cultural Heritage field because specific training is a necessary step.

From a user perspective, data preparation can also be a critical issue. The tool VGG Image Annotator (VIA) has been used for the annotation of the bounding box of the architectural heritage. VIA is a simple and standalone manual annotation tool for images, audio and video that allows the description of spatial regions in images or video frames. These manual annotations can be exported to plain text data formats such as

JSON and CSV and therefore are ready for further processing by other software tools (Dutta and Zisserman, 2019).

The file with the bounding box coordinates was used to prepare the dataset according to the requirements of Luminoth. After that, a configuration file has to be created specifying some necessary information, such as a run name, the location of the dataset and the model to use to train the network.

Luminoth also allows users to select the hyperparameters of the training, i.e., the parameters whose values are used to control the learning process. The selection is performed by manually customizing the training configuration file. Tuning the hyperparameters can be crucial to optimally solve the machine learning problem, e.g., in terms of convergence, stability and performance of training and inference phases. All in all, the default values provided by Luminoth mostly demonstrated to be effective in terms of all of the objectives. In particular, the momentum optimizer (Ruder, 2017) was adequate to reach convergence. As for the learning rate tuning, the default values were usually adequate but for some SSD-based training runs, some convergence issues arose, and these problems were addressed by modifying the learning rate value from the default (0.0003) to lower values (e.g., 0.00006). Luminoth also integrates an automatic data augmentation mechanism and it was helpful to increase the entropy of data used in the training. As concerns the number of epochs to be used during the training, it was manually selected to ensure a complete training convergence. An early stopping mechanism might be possible but it was not attempted so far to always get the best results for the considered test cases. Finally, in Chapter 4 of this work, an analysis of computing time performances comparing different generations of hardware is provided to complete the picture also from that point of view.

3.2.3 Evaluation Metrics of Neural Networks

As introduced, the NNs potentially improve the efficiency of the first part of the photogrammetric workflow. However, it is necessary to evaluate this performance more objectively, directly addressing also the efficiency and reliability of the algorithm in reducing the effort of the final user activity. According to these considerations, two different types of metrics evaluation of the network are considered. In the first type, the efficiency of the performance of NN is evaluated using standard metrics based on images or frames evaluated separately, while in the second type of evaluation the metrics are more closely related to the final tasks of the network, i.e. discovering the time intervals where a selected object appears in a video minimizing the human time required to manually analyse the movies. According to the frame-based approach, standard conventions can be followed: given a dataset of images, it is first defined P (N) as the number of images in which the object is present (not present), respectively. During the real-world inference phase, these values are not known, and the network output is P’ (N’) that represents the images in which the network has found (not found) the object. When performing the object detection inference, a probability of the presence of the object is typically returned. Therefore, in order to get the P’ and N’ values, it is necessary to define a probability threshold which is the minimum probability to be returned to consider the object as found. In order to validate the network performances, a test phase in which P and N are known is taken into consideration so that it is possible to categorize the images according to 4 statuses: True Positive (TP, image in both P and P '), True Negative (TN, image both in N and in N '), False Positive (FP, image in P' but not in P) and False Negative (FN, image in N 'but not in N). Obviously, T=TP+FN and N=TN+FP. Such quantities can be combined to define meaningful parameters. In particular, a typical indicator is the accuracy, calculated as:

$$\mathrm{AC} = \frac{\mathrm{TP} + \mathrm{T}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

Two other typical indicators are:

Sensitivity (SN): defined as the number of correct positive predictions divided by the total number of positives:

$$\mathrm{SN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

Specificity (SP): defined as the number of correct negative predictions divided by the total number of negatives:

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{3}$$

Considering a set of test images composed only of positive (negative) images, it is clear that TN=FP=0 (TP=FN=0) and the accuracy exactly corresponds to the sensitivity (specificity). As detailed in the next section, considering positive-only (or negative-only) sets is very useful during the network training and validation phase to evaluate the different capabilities of the network.

The indicators above are useful because they can work with both images and video frames, allowing fine-grained comparisons, and are especially useful to assess the quality of the network during the training phase. However, considering the usage of the network in a real-world context, where a certain object has to be detected from a large number of video archives, the authors believe that a set of metrics based on the time intervals is more suitable to summarize the advantages of using NNs in preference to the manual alternative. Referring these measures to the intervals is more natural if we consider that once a time interval with the searched object is found also with a single-frame, it is easy to identify the correct time set in which the monument appeared simply by going back or forward in the video. The proposed metrics are, therefore:

- Discovery Rate (DR): calculated as the number of the intervals correctly predicted by the network divided by the total number of the true intervals:

$$\mathrm{DR} = \frac{\text{TP (n. correctly predicted intervals)}}{\text{P (n. correctly true intervals)}} \tag{4}$$

A time interval where the searched object appears is considered “correctly predicted” if there is a predicted interval that overlaps with the true interval for at least 1 second of video. This discovery rate is somehow related to the sensitivity: indeed, it indicates a measure of the correct positive predictions over the total number of positive cases. This metric describes an issue that is very important for the user, i.e. the capacity of the network to detect monuments avoiding loss of information.

- Time save Rate (TSR): calculated as the total time length of the video divided by the sum of the times of the measured intervals:

$$\mathrm{TSR} = \frac{\text{Total video length}}{\text{Sum. times of predicted intervals}} \tag{5}$$

This parameter is somehow related to the specificity and indicates how much time the operator would save in his manual work of watching videos if the automatism of the network is used. This parameter clearly depends also on the type of videos used for testing. The vast majority of the results presented in this thesis are based on videos which contain at least one occurrence of the searched object, such a circumstance artificially limits the measured time save rate. To circumvent the dependency on the archive source, another time save related parameter is defined.

- Time save Efficiency (TSE): as said, the time save rate indicates the quantitative advantage for the end-user who adopts NN. It would be interesting to compare the time save rate with the ideal time save rate, which is the minimum time save rate knowing in advance the true intervals to watch. The ratio between the measured time save and the ideal time save rate is the time save efficiency:

$$\begin{aligned}\mathrm{TSE} = \frac{\text{TSR measured}}{\text{TSR ideal}} &= \frac{\text{Total video length}\big/\text{Sum. times of predicted intervals}}{\text{Total video length}\big/\text{Sum. times of true intervals}} \\ &= \frac{\text{Sum. times of true intervals}}{\text{Sum. times}}\end{aligned} \tag{6}$$

The time save efficiency is less prone to bias due to the type of archive source. The value of this efficiency is typically reduced when dealing with false positives, i.e. when some measured intervals do not correspond to true intervals. However, it is worth noting that this efficiency can also reach values greater than 1 and this can happen when not all the true intervals are correctly found by the measured intervals. In this scenario, despite the false-positive intervals, it is possible that the total time of the measured intervals becomes shorter than the time of true intervals. However, this is a clear symptom of a poor discovery rate.

In general, discovery-rate, time save rate, and time save efficiency should all be maximized to improve training but, when it is not possible, the choice of one metric or another is a matter of usage context. Because of the type of use of the network within the photogrammetric pipeline, two extreme cases of use are considered:

1. In the first situation in which the videos selected by the DL are then manually watched to decide which are the most suitable for photogrammetry, it is ideal to maximize the discovery rate to avoid losing useful information.

2. In the second situation in which the pipeline is managed more automatically, it is instead preferable to maximize the time save parameters to prevent incorrect images from entering the subsequent processing.

3.3 Camera motions identification for photogrammetry

In the second phase of the workflow, the frames that can be processed by photogrammetry are selected from all images detected by the Neural Network. The selection is performed according to the camera motions in the video scene. As shown in Chapter 2, the tilting and trucking camera motions have demonstrated to be more effective to perform this kind of frame selection since only images taken from multiple angles in the same scene are suitable for photogrammetry.

The algorithm for determining the camera motions from the results of object recognition by the Neural Network is shown in Figure 5 and detailed below, while the python script is reported in Appendix A.

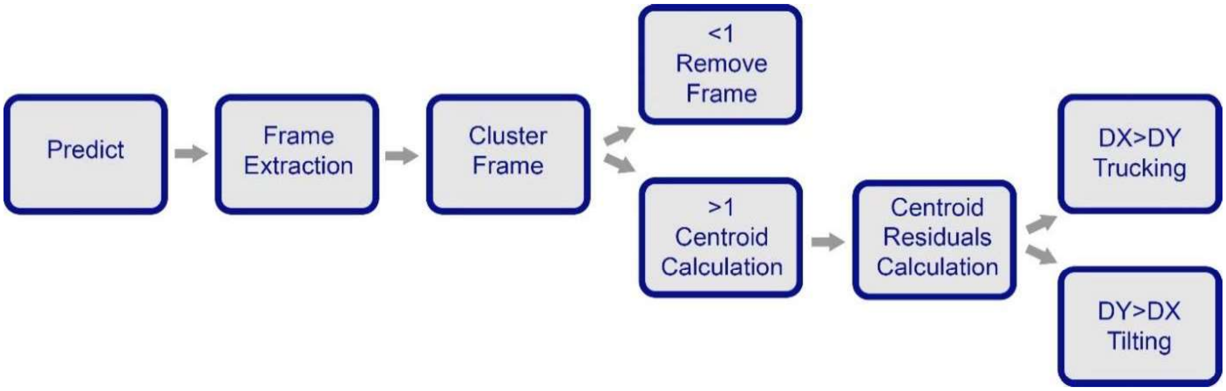


Figure 5. Workflow of the second step of the proposed match-moving method: the tilting and trucking camera motions identification suitable for photogrammetry.

1. Predict: the object detection algorithm ends with the predict step, which lists the frames where the searched monument appears. Each frame is uniquely identified by the name of the video and the time appearance within the video. The object detection also returns the coordinates of the position of the bounding boxes in the frame and the probability score of the presence of the monument in the video. Based on a predefined probability threshold, positive frames with a score below the threshold are removed from the list of predicted objects.

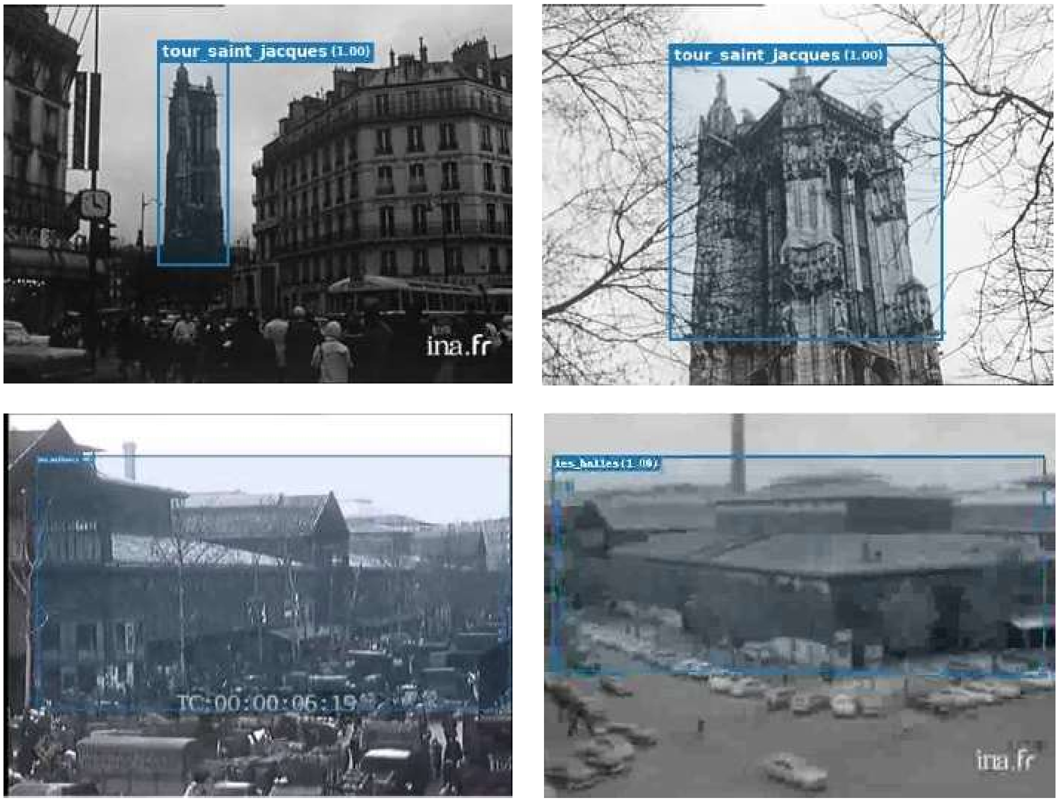


Figure 6. An example of the frame selected by the predict step of the algorithm.

2. Image extraction: from the results of the previous step, the frames are extracted from the video and saved as separate images. It should be noted that the photogrammetric method may start from this set of images, but a high failure rate can be expected when simultaneously using images extracted from different videos (with different characteristics and qualities), different scenes and different camera movements. This usually requires manual intervention and decisions to achieve a successful final photogrammetric reconstruction. In order to automate a successful procedure, a further elaboration of extracted frames is proposed.

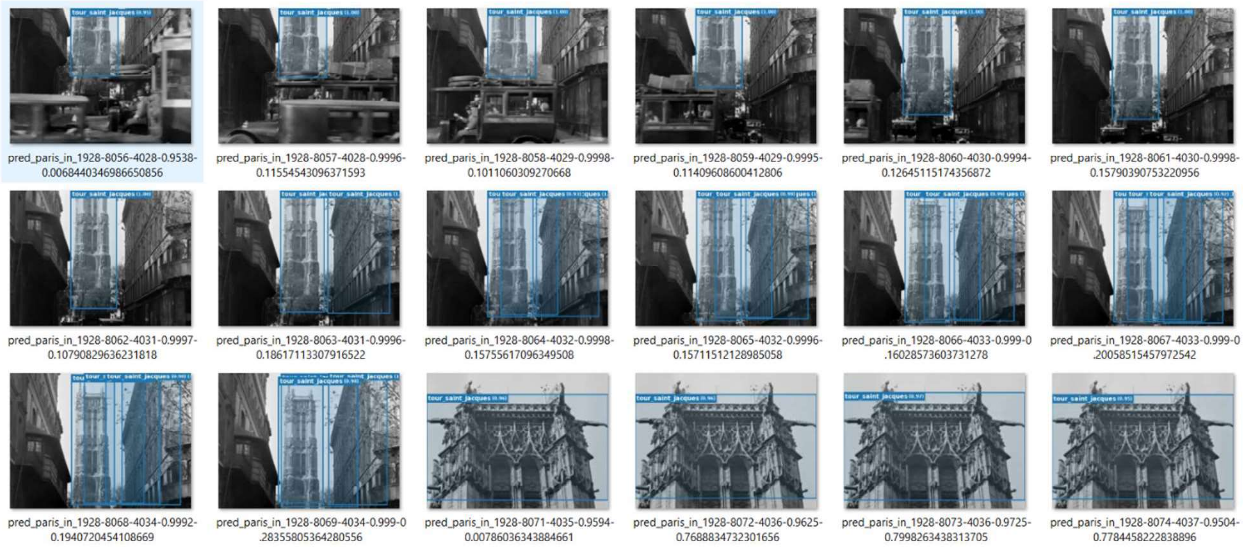
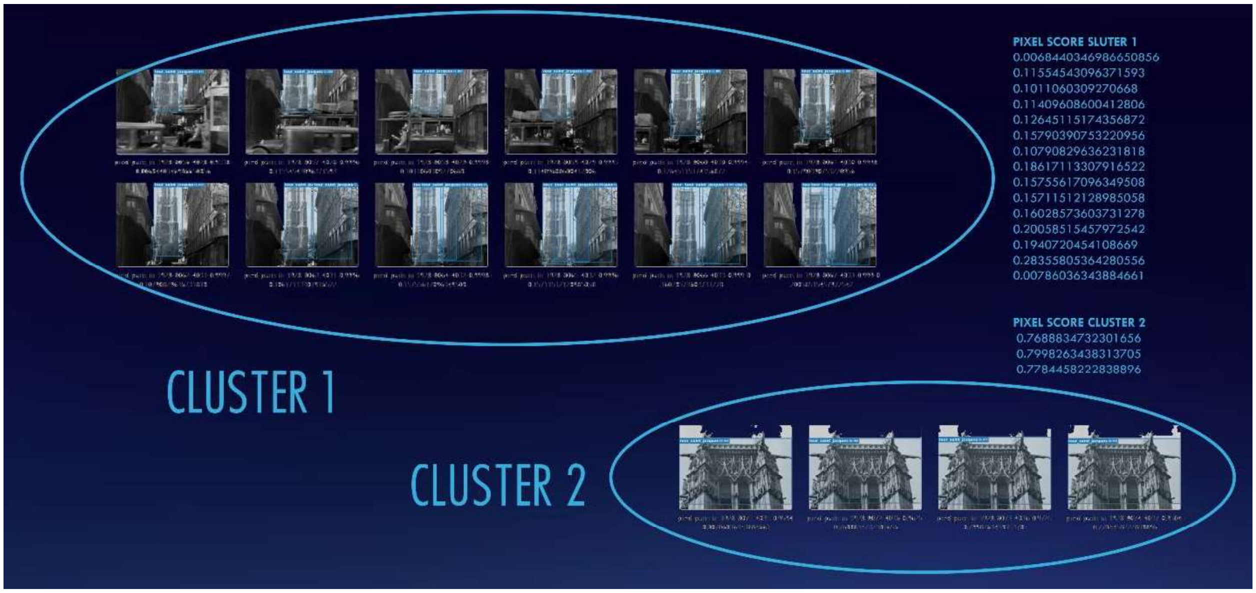

pred_paris_in_1928-8056-4028-0.9538-0.0068440346986650856
pred_paris_in_1928-8057-4028-0.9996-0.115545430963715 93
pred_paris_in_1928-8058-4029-0.9998-0.1011060309270668
pred_paris_in_1928-8059-4029-0.9995-0.11409608600412806
pred_paris_in_1928-8060-4030-0.9994-0.12645115174356872
pred_paris_in_1928-8061-4030-0.9998-0.15790390753220956
pred_paris_in_1928-8062-4031-0.9997-0.10790829636231818
pred_paris_in_1928-8063-4031-0.9996-0.18617113307916522
pred_paris_in_1928-8064-4032-0.9998-0.15755617096349508
pred_paris_in_1928-8065-4032-0.9996-0.15711512128985058
pred_paris_in_1928-8066-4033-0.999-0.16028573603731278
pred_paris_in_1928-8067-4033-0.999-0.20058515457972542
pred_paris_in_1928-8068-4034-0.9992-0.1940720454108669
pred_paris_in_1928-8069-4034-0.999-0.28355805364280556
pred_paris_in_1928-8071-4035-0.9594-0.00786036343884661
pred_paris_in_1928-8072-4036-0.9625-0.7688834732301656
pred_paris_in_1928-8073-4036-0.9725-0.7998263438313705
pred_paris_in_1928-8074-4037-0.9504-0.7784458222838896

Figure 7. An example of the frame extracted by the image extraction step of the algorithm.



PIXEL SCORE SLUTER 1
0.0068440346986650856
0.11554543096371593
0.1011060309270668
0.11409608600412806
0.12645115174356872
0.15790390753220956
0.10790829636231818
0.18617113307916522
0.15755617096349508
0.15711512128985058
0.16028573603731278
0.20058515457972542
0.1940720454108669
0.28355805364280556
0.00786036343884661

PIXEL SCORE CLUSTER 2
0.7688834732301656
0.7998263438313705
0.7784458222838896

Figure 8. An example of the cluster of frame created by the frame clustering step of the algorithm.

3. Frame clustering: the extracted frames are grouped in two different splitting criteria, both aim to detect a change of the scene in the video. In particular, a new frame cluster is created if at least one of the following criteria is met:

a. The first one is time-based: frames that are consecutive in the time-line belong to the same group.

b. The second one relies on a structural similarity comparison (Wang et al., 2004; Avanaki 2009). Looping over the selected frames, a similarity score between 0 and 1 is evaluated for each frame compared to the previous one. If the score is less than a predefined threshold Ts, a new group is created for the analysed frame.

4. Cluster cleaning: since the intention is to perform photogrammetry on each cluster of frames, clusters with only one image are marked as invalid. Using frames belonging to the same cluster of intervals leads to a higher success rate of the photogrammetric process. In fact, putting together single frames taken at different time intervals or from different videos can certainly help to recover more information about the lost heritage, but at the same time there is a higher risk that the photogrammetric process fails because they are dated in different historical time periods.

5. Bounding boxes centroid calculation: for each frame in a valid cluster, the centroid [xc,yc] of the bounding box is computed in order to analyse the position change of the object.

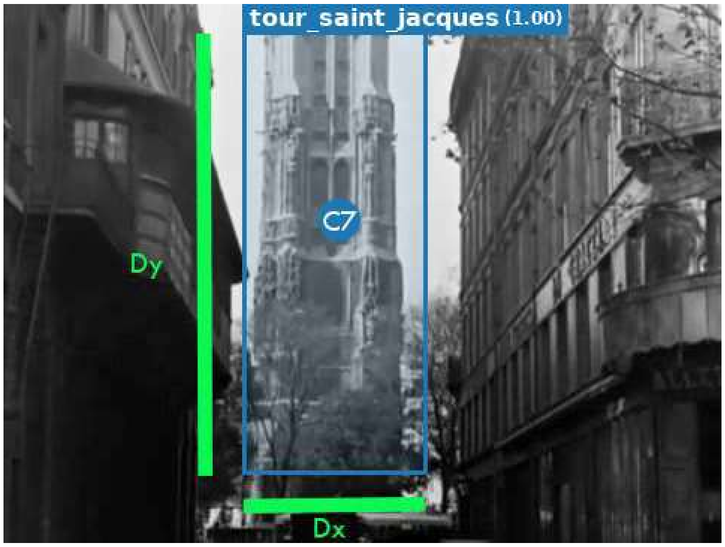


Figure 9. An example of the centroid calculation by the algorithm.

6. Centroid residuals calculation: for each frame cluster, the cumulative residuals between the frame centroids are evaluated using the first and last frame centroids of the cluster:

$$Dx = xc(\text{last frame}) - xc(\text{first frame})$$
$$Dy = yc(\text{last frame}) - yc(\text{first frame})$$

7. Camera motion estimation: These residuals are used to guess the camera motion. Clearly, the detection of centroid movements is not sufficient to accurately

evaluate the camera motion. However, simple assumptions can lead to satisfactory results for the success of the entire proposed procedure. In particular, when |Dx| > |Dy| the trucking camera motion is expected while camera tilting corresponds to the |Dx| < |Dy| case. This simple assumption may lead to wrong results for common cases. If both Dx and Dy are very small, this could correspond to a fixed camera poorly anchored to the terrain (a tripod was not used for most historical film footage). On the other hand, if Dx and Dy are not small but close to each other, camera motion guessing is more questionable. For these reasons, the proposed algorithm uses can be summarized in the following Figure 9, where T1 and T2 are two thresholds. In the end, the devised algorithm allows the user to distinguish among four camera motion categories, namely: "steady or oscillating camera", "camera trucking", "camera tilting" and "cannot determine camera motion". This simple categorization is however useful in view of the final purpose of the algorithm that is to detect frame clusters useful for photogrammetry.

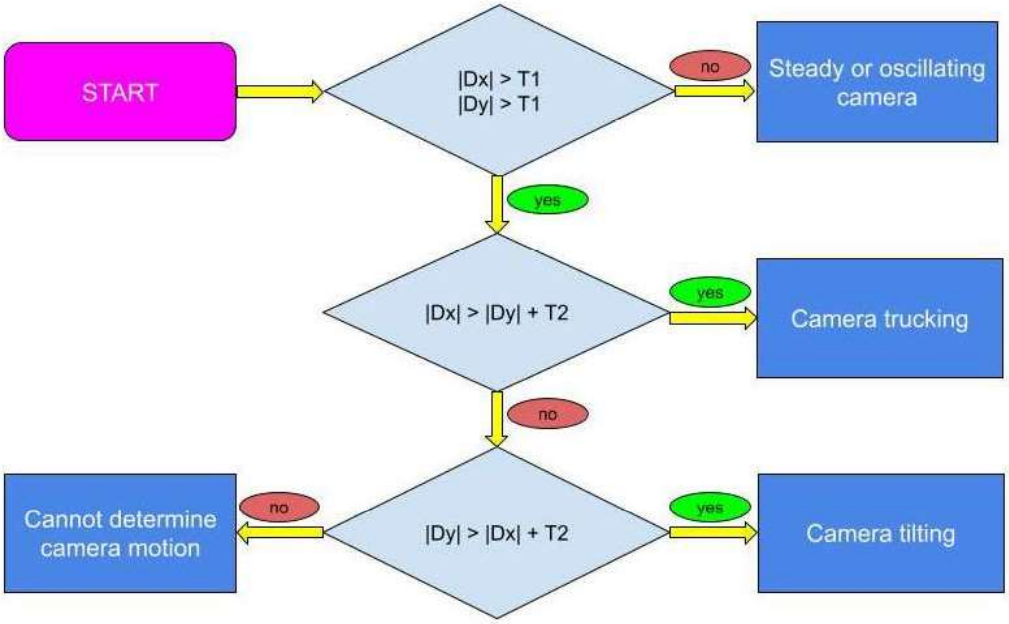


Figure 10. Workflow of the camera motion estimation step of the algorithm.

As discussed above, some of the steps described lead to automatic data filtering. Since filtering is an automatic process, it can lead to the elimination of data that is potentially useful for the final process. Decreasing the introduced thresholds limits the loss of data, but puts at risk the success of the automatic procedure, at least in some cases. Depending on the context of use, it is necessary to evaluate the choices to make. Overall, the choice to go towards an automatic procedure leads to a significant improvement in the efficiency of the process, both in terms of time and in terms of simplicity of execution. In these cases, a certain penalty in terms of the ability to exploit the single data should be acceptable.

References

Andrianaivo, L. N., D'Autilia, R., and Palma, V., 2019. Architecture recognition by means of convolutional neural networks, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W15, 77–84, https://doi.org/10.5194/isprs-archives-XLII-2-W15-77-2019, 2019.

Avanaki, A.N., 2009. Exact global histogram specification optimized for structural similarity. Optical Review, 16, 613-621, arXiv.org/0901.0065, doi.org/10.1007/s10043-009-0119-z.

Belhi, A., Bouras, A., Foufou, S., 2018. Leveraging Known data for missing label prediction in Cultural Heritage context. Appl. Sci. 2018, 8, 1768, doi:10.3390/app8101768.

Bhargav, S., van Noord, N., Kamps, J., 2019. Deep learning as a tool for early cinema analysis. In Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC '19), 21 October 2019. Association for Computing Machinery, New York, NY, USA; pp. 61–68, doi:10.1145/3347317.3357240.

Boujou, Vicon Motion Systems Ltd UK. http://www.vicon.com/software/boujou/ (1 May 2020).

Chapel, M.N., Bouwmans, T., 2020. Moving Objects Detection with a Moving Camera: A Comprehensive Review. Computer Vision and Pattern Recognition, arxiv.org/abs/2001.05238.

Condell, J., Moore, G., 2006. Software and Methods for Motion Capture and Tracking in Animation. Proceedings of the 2006 International Conference on Computer Graphics & Virtual Reality, CGVR 2006, CSREA Press 2006, 3-9.

Condorelli, F., Rinaudo, F., Salvadore, F., Tagliaventi, S., 2020. A match-moving method combining AI and SfM algorithms in historical film footage, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B2-2020, 813–820, https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-813-2020, 2020.

Condorelli, F., Rinaudo, F., Salvadore, F., Tagliaventi, S., 2020. A Neural Networks Approach to Detecting Lost Heritage in Historical Video, ISPRS Int. J. Geo-Inf. 2020, 9(5), 297; https://doi.org/10.3390/ijgi9050297.

Condorelli, F. and Rinaudo, F., 2019. Processing historical film footage with Photogrammetry and Machine Learning for Cultural Heritage documentation. In MM '19: 2019 ACM Multimedia Conference Proceedings, October 2019, Nice, France. ACM, NY, USA. 8 pages. https://doi.org/10.1145/3347317.3357248.

Condorelli, F., Rinaudo, F., Salvadore, F., and Tagliaventi, S., 2019. Architectural Heritage recognition in historical film footage using Neural Networks, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W15, 343–350, https://doi.org/10.5194/isprs-archives-XLII-2-W15-343-2019, 2019.

Dağlar, O., Tong, T., 2011. A Method on Using Video in Architectural Design Process: Matchmoving. Respecting Fragile Places: 29th eCAADe Conference Proceedings, 339-348.

Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A., 2012. What Makes Paris Look like Paris?. ACM Transactions on Graphics, Association for Computing Machinery, 2012, 31 (4). ⟨hal-01053876⟩

Dong, Z., Zhang, G., Jia, J., Bao, H., 2009. Keyframe-Based Real-Time Camera Tracking. IEEE International Conference on Computer Vision (ICCV), 118,1538 - 1545 doi.org.0.1109/ICCV.2009.5459273.

Doulamis, A., 2018. Automatic 3D Reconstruction From Unstructured Videos Combining Video Summarization and Structure From Motion. Front. ICT 5:29. doi: 10.3389/fict.2018.00029.

Dutta, A., Zisserman, A., 2019. The via annotation software for images, audio and video. In Proceedings of the 27th ACM International Conference on Multimedia (MM '19), Nice, France, 21–25 October 2019; ACM: New York, NY, USA; 276–2279, doi:10.1145/3343031.3350535.

Fiorucci, A., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., Jame, S., 2020. Machine Learning for Cultural Heritage: A Survey. Pattern Recognition Letters 133 (2020) 102–108.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587, doi:10.1109/CVPR.2014.81.

Gominski, D., Poreba, M., Gouet-Brunet, V., Chen, L., 2019. Challenging deep image descriptors for retrieval in heterogeneous iconographic collections. In Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC '19), 21 October 2019. Association for Computing Machinery, New York, NY, USA; pp. 31–38, doi:10.1145/3347317.3357246.

Haji, K., Sharif, A. P., Rabbani, I. A., 2016. An Overview of Matchmoving using Structure from Motion Methods. Proceedings of the 2016 Symposium on Digital Production, 45-54, doi.10.1145/2947688.2947697.

Ingwer, P., Gassen, F., Püst, S., Duhn, M., Schälicke, M., Müller, K., Ruhm, H., Rettig, J., Hasche, E., Fischer, E., Creutzburg, R., 2015. Practical usefulness of structure from motion (SfM) point clouds obtained from different consumer cameras. Proc. SPIE 9411, Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2015, 941102, doi: 10.1117/12.2074892.

Khademi, S., Shi, X., Mager, T., Siebes, R., Hein, C., Boer, V.D., & Gemert, J.C., 2018. Sight-Seeing in the Eyes of Deep Neural Networks. 2018 IEEE 14th International Conference on e-Science (e-Science), 407-408.

Krizhevsky, A.; Sutskever, I.; Hinton, G., 2012. Imagenet classification with deep convolutional neural networks. NIPS 2012, 1, 1097–1105, doi:10.1145/3065386.

Lee, B., Park, J., Young Sung, M., 2006. Vision-Based Real-Time Camera Matchmoving with a Known Marker. Proceedings of the 5th international conference on Entertainment Computing, 193-204, doi.10.1007/11872320_23.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., 2016. SSD: Single shot multibox detector. Comput. Vis. Eccv Lect. Notes Comput. Sci. 2016, 9905, 21–37, doi:10.1007/978-3-319-46448-0_2.

Llamas, J., Lerones, P.M., Medina, R., Zalama, E., Gómez-García-Bermejo, J., 2017. Classification of architectural heritage images using deep learning techniques. Appl. Sci. 2017, 7, 992, doi:10.3390/app7100992.

Lourakis, M., Argyros, A., 2005. Camera Matchmoving in Unprepared, Unknown Environments. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, doi.org10.1109/CVPR.2005.96.

Luminoth Open Source Computer Vision Toolkit. Available online: https://github.com/tryolabs/luminoth (accessed on 9 March 2020).

Luo, A., Chen, S., Tseng, K., 2016. A real-time camera match-moving method for virtual-real synthesis image composition using temporal depth fusion. International Conference on Optoelectronics and Image Processing (ICOIP), 35-39, doi.org. 10.1109/OPTIP.2016.7528515.

Montoya Obeso, A., Benois-Pineau, J., Saraí García Vázquez, M., Ramírez Acosta, A., 2019. A. organizing Cultural Heritage with deep features. In Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC '19), 21 October 2019, Association for Computing Machinery, New York, NY, USA; pp. 55–59, doi:10.1145/3347317.3357244.

Parekh, H., Thakore, D., Jaliya, K., 2014. A Survey on Object Detection and Tracking Methods. International Journal of Innovative Research in Computer and Communication, 2(2), 2979-2978.

Picard, D., Gosselin, P.H., Gaspard, M.C., 2105. Challenges in content-based image indexing of Cultural Heritage collections. IEEE Signal Process. Mag. Inst. Electr. Electron. Eng. 2015, 32, 95, doi:10.1109/MSP.2015.2409557.

Radenović, F., Tolias, G., Chum, O., 2016. CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples. Eur. Conf. Comput. Vis. ECCV 2016, 1–17, doi:10.1007/978-3-319-46448-0_1.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 2016, 39, 1–13, doi:10.1109/TPAMI.2016.2577031.

Ruder, S., 2017. An overview of gradient descent optimization algorithms. arXiv 2017, arXiv:1609.04747v2.

Saini, A., Gupta, T., Kumar, G., Kumar Gupta, A., Panwar, M., Mittal, A., 2017. Image based indian monument recognition using convoluted neural networks. Int. Conf. Big Data IOT Data Sci. 2017, 1–5, doi:10.1109/BID.2017.8336587.

Shi, X., Khademi, S., van Gemert, J., 2019. Deep visual city recognition visualization. arxiv 2019, 1–6, arxiv:1905.01932.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. arxiv 2015, 1–14, arxiv:1409.1556.

SynthEyes, Andersson Technologies LLC. https://www.ssontech.com/ (1 May 2020).

Szegedy, C., Liu, X., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. arxiv 2015, 1–12, arxiv:1409.4842.

TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Available online: http://tensorflow.org (accessed on 9 March 2020).

Toda, T., Masuyama, G., Umeda, K., 2016. Detecting Moving Objects Using Optical Flow with a Moving Stereo Camera. MOBIQUITOUS 2016: Adjunct Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing Networking and Services, 35–40, doi.org/10.1145/3004010.3004016.

Tolias, G., Sicre, R., Jégou, H., 2016. Particular object retrieval with integral max-pooling of CNN activations. arxiv 2016, 1–12, arxiv: 1511.05879.

Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., 2004. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13, 600-612, doi.org/10.1109/TIP.2003.81986.

Yasser, A.M., Clawson, K., Bowerman, C. , 2017. Saving Cultural Heritage with digital make-believe: machine learning and digital techniques to the rescue. In Proceedings of the Electronic Visualisation and the Arts, London, UK, 11–13 July 2017; pp. 1–5, doi:10.14236/ewic/HCI2017.97.

Zhang, G., Dong, Z., Jia, J., Wan, L., Wong, T., Bao, H., 2009. Refilming with Depth-Inferred Videos. IEEE Transactions on Visualization and Computer Graphics, 15(5), 828-840.

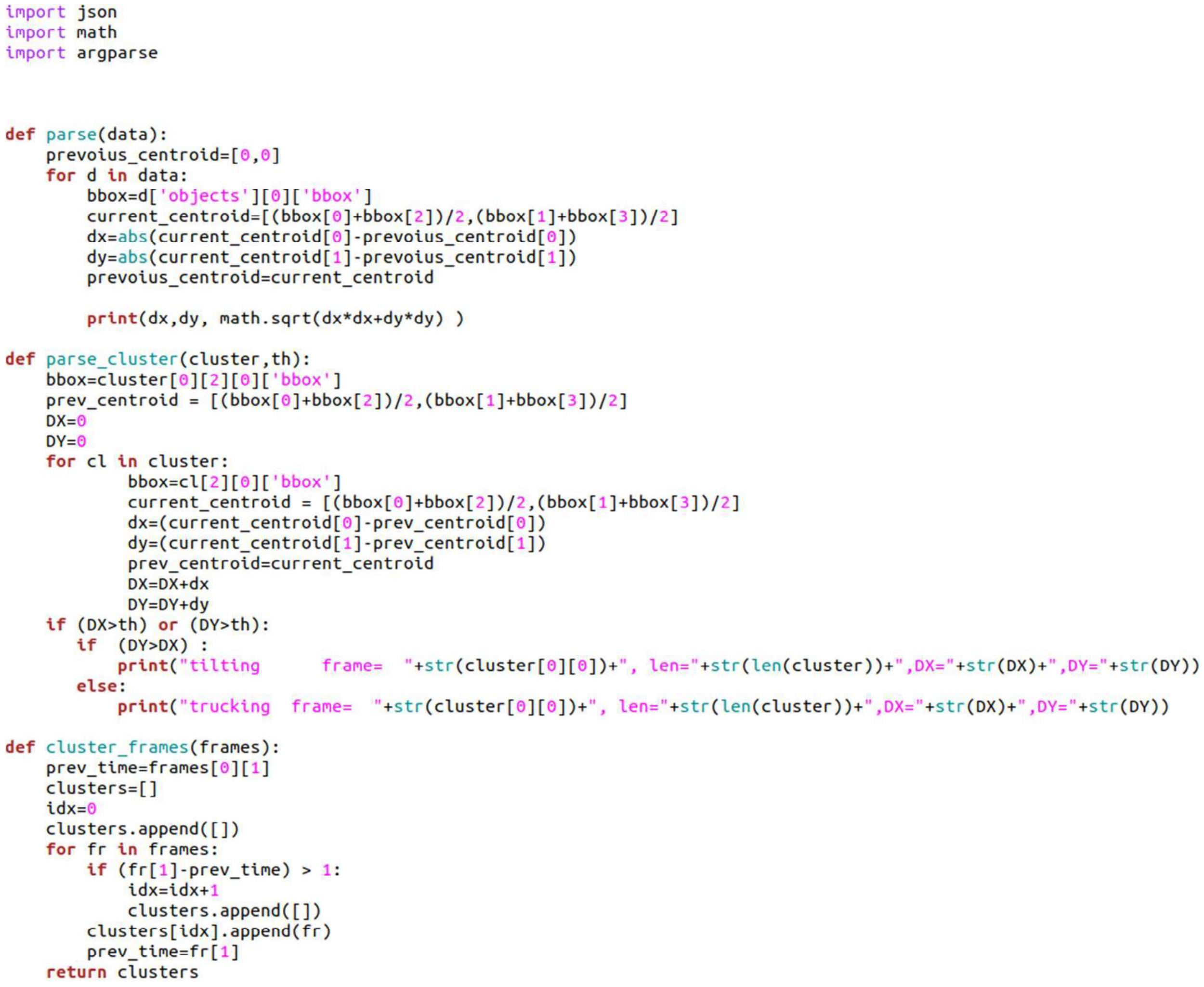
Zhang, G., Liu, H., Dong, Z., Jia, J., Wong, T., Bao, H., 2016. Efficient Non-Consecutive Feature Tracking for Robust Structure-From-Motion. IEEE Transactions on image processing, 25(12), 422-435, doi.org.10.1109/TIP.2016.2607425.

Zhang, G., Qin, X., Hua, W., Wong, T., Heng P., Bao, H., 2007. Robust Metric Reconstruction from Challenging Video Sequences. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1-8.

Zhang, G., Dong, Z., Jia, J., Wan, L., Wong, T., Bao, H., 2009. ACTS: Automatic Camera Tracking System. http://www.zjucvg.net/acts/acts.html (1 May 2020).

Appendix A

Algorithm of camera motion identification for photogrammetry



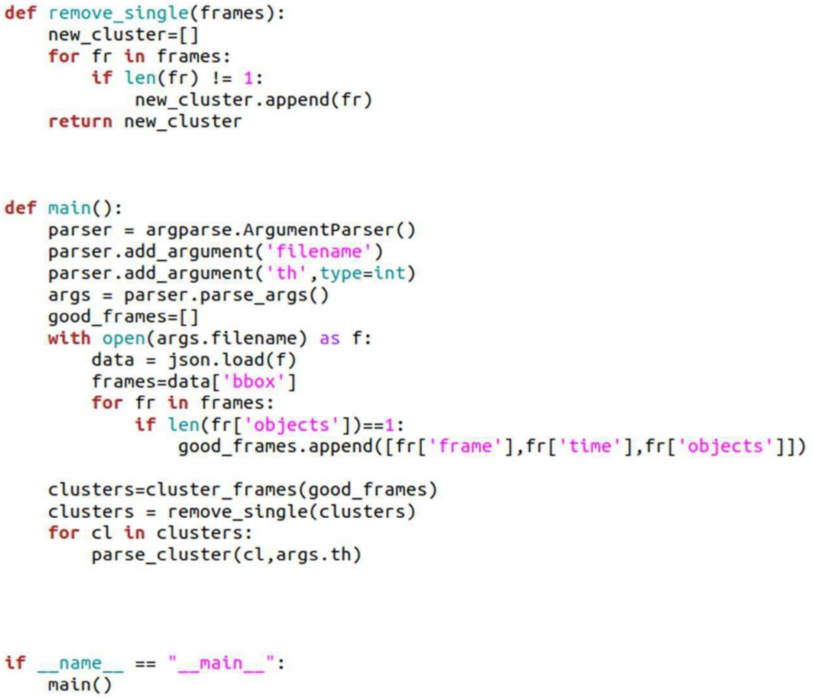
```
import json
import math
import argparse


def parse(data):
    prevoius_centroid=[0,0]
    for d in data:
        bbox=d['objects'][0]['bbox']
        current_centroid=[(bbox[0]+bbox[2])/2,(bbox[1]+bbox[3])/2]
        dx=abs(current_centroid[0]-prevoius_centroid[0])
        dy=abs(current_centroid[1]-prevoius_centroid[1])
        prevoius_centroid=current_centroid

        print(dx,dy, math.sqrt(dx*dx+dy*dy) )

def parse_cluster(cluster,th):
    bbox=cluster[0][2][0]['bbox']
    prev_centroid = [(bbox[0]+bbox[2])/2,(bbox[1]+bbox[3])/2]
    DX=0
    DY=0
    for cl in cluster:
            bbox=cl[2][0]['bbox']
            current_centroid = [(bbox[0]+bbox[2])/2,(bbox[1]+bbox[3])/2]
            dx=(current_centroid[0]-prev_centroid[0])
            dy=(current_centroid[1]-prev_centroid[1])
            prev_centroid=current_centroid
            DX=DX+dx
            DY=DY+dy
    if (DX>th) or (DY>th):
       if  (DY>DX) :
           print("tilting      frame=  "+str(cluster[0][0])+", len="+str(len(cluster))+",DX="+str(DX)+",DY="+str(DY))
       else:
           print("trucking  frame=  "+str(cluster[0][0])+", len="+str(len(cluster))+",DX="+str(DX)+",DY="+str(DY))

def cluster_frames(frames):
    prev_time=frames[0][1]
    clusters=[]
    idx=0
    clusters.append([])
    for fr in frames:
        if (fr[1]-prev_time) > 1:
            idx=idx+1
            clusters.append([])
        clusters[idx].append(fr)
        prev_time=fr[1]
    return clusters
```



```
def remove_single(frames):
    new_cluster=[]
    for fr in frames:
        if len(fr) != 1:
            new_cluster.append(fr)
    return new_cluster



def main():
    parser = argparse.ArgumentParser()
    parser.add_argument('filename')
    parser.add_argument('th',type=int)
    args = parser.parse_args()
    good_frames=[]
    with open(args.filename) as f:
        data = json.load(f)
        frames=data['bbox']
        for fr in frames:
            if len(fr['objects'])==1:
                good_frames.append([fr['frame'],fr['time'],fr['objects']])

    clusters=cluster_frames(good_frames)
    clusters = remove_single(clusters)
    for cl in clusters:
        parse_cluster(cl,args.th)



if __name__ == "__main__":
    main()
```

Chapter 4

Applications on case studies and results

In this section the case studies chosen to test the workflow for the processing of historical images proposed in this dissertation are described and the main results are discussed.

The approach used is to apply the same methodology to different scales, referring to the buildings and groups of buildings scales.

In particular, in the first part of this section the method was tested applied the entire workflow on two of the different situations in which the heritage could be, in this case they are lost and transformed monuments. Thanks to the participation at the "Summer School: Cities Cultural Heritage and Digital Humanities" organized by prof. Tamborrino in 2018, monuments in Paris were studied and different historical sources analysed and made available for the research. Among these materials, the following monuments were selected as an application in this dissertation: the Tour Saint Jacques (Figure 3), a UNESCO World Heritage Site that is now existing, and the pavilions of Les Halles (Figure 4) by Baltard that no longer exist since they were destroyed in 1971.

Starting from the application of Artificial Intelligence on film footage to select suitable material for photogrammetry, the test continues with the 3D reconstruction and the metric quality assessment of the models obtained to compare and to discuss the results according to different type of material and algorithms used during the processing.

In the second part of this section the discussion of results is concentrated on the photogrammetric part of the workflow, testing it on both film footage and historical photograph of other two case studies, representing ephemeral architecture and restored heritage. The choice of this application derived from other two experiences made

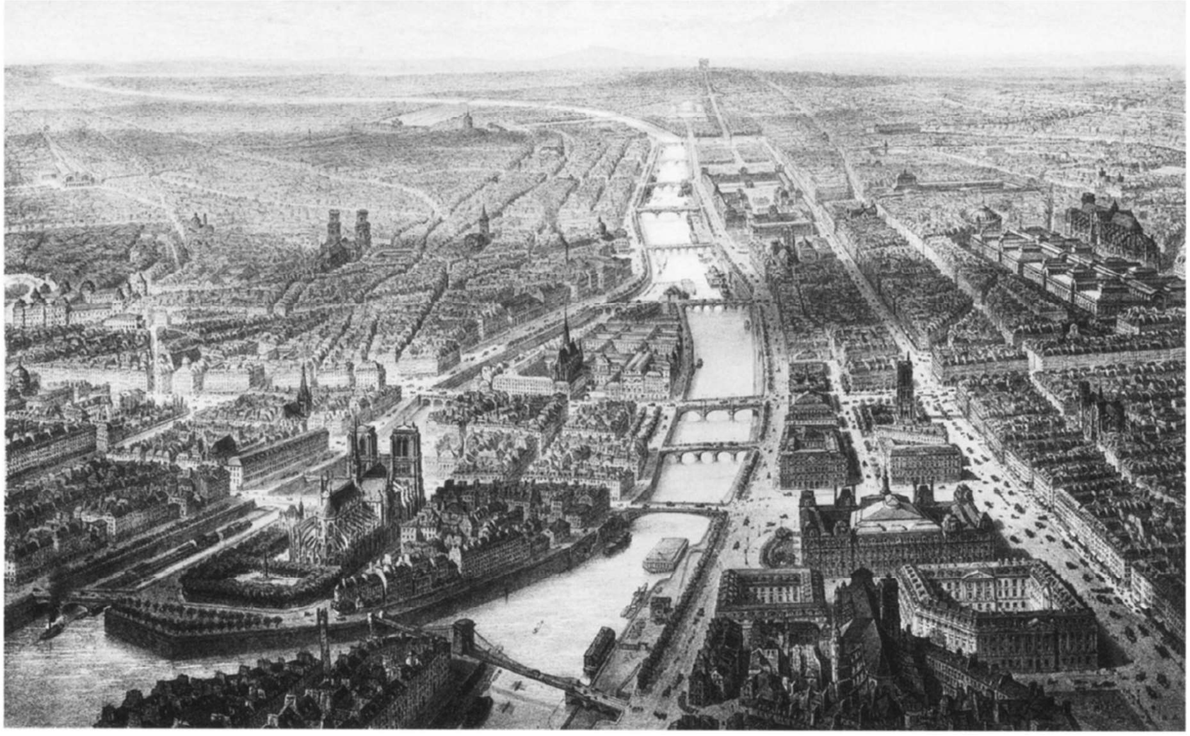


Figure 1. Paris in 1860, view of the Saint-Gervais Quarter, showing the location of the Tour Saint-Jacques at lower right, on the Seine's Right Bank and the Pavilions of Les Halles. From Philippe Benoist, Paris dans sa splendeur (Paris, 1861).

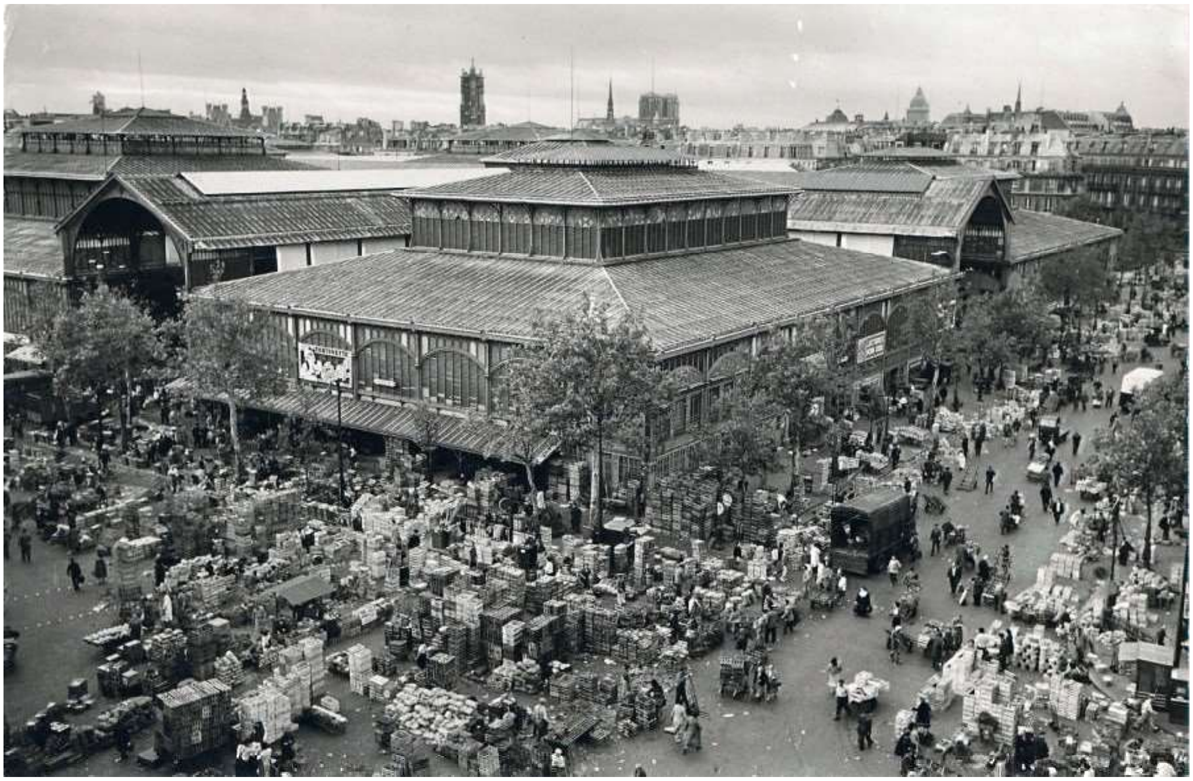
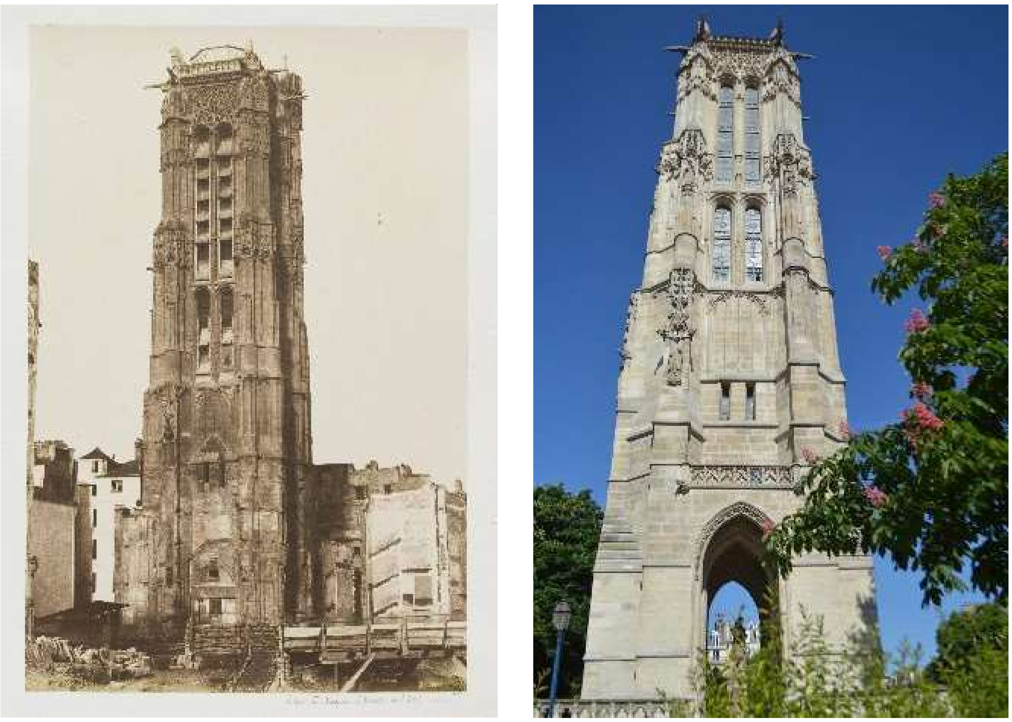
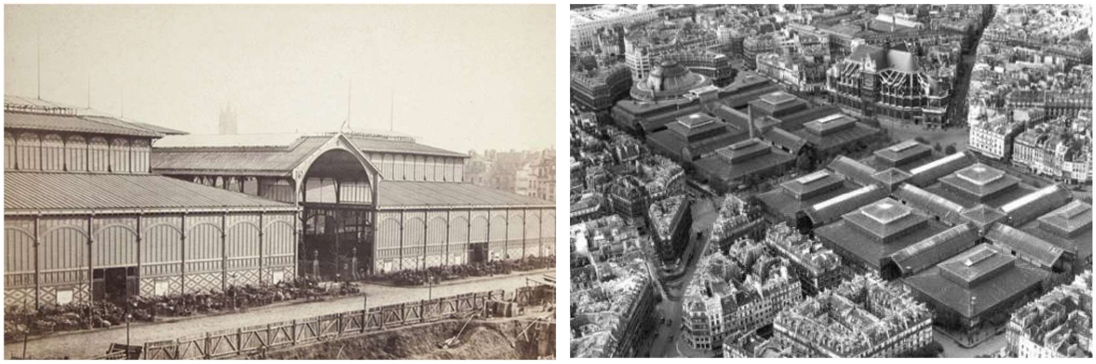


Figure 2. View of Les Halles and the Tour Saint Jacques,1969, Archives de Paris.



(a) (b)

Figure 3. Tour Saint-Jacques la Boucherie (1508-22), Paris. (a) Henri Jean-Louis Le Secq, 1853, Musée Carnavalet. (b) Francesca Condorelli (author), 2019.



(a) (b)

Figure 4. Les Halles centrales by the architect Victor Baltard. (a) Pavilions 5 and 6, Charles Marville, 1855, Musée Carnavalet / Roger-Viollet. (b) Top view of Les Halles, Roger Henrard, 1952, Musée Carnavalet/ Roger-Viollet.

during the research activity: the test of I-Media-Cities platform (see Chapter 1), with partner CINECA research centre and the Cinema Museum of Torino, that made available the film footage concerning International Exposition in Torino, and the collaboration with Tokyo Institute of Technology during the visiting research period that made available material concerning archival images of Japanese heritage building and wall paintings in Cappadocia. The results were used to validate the method proposed in this work.

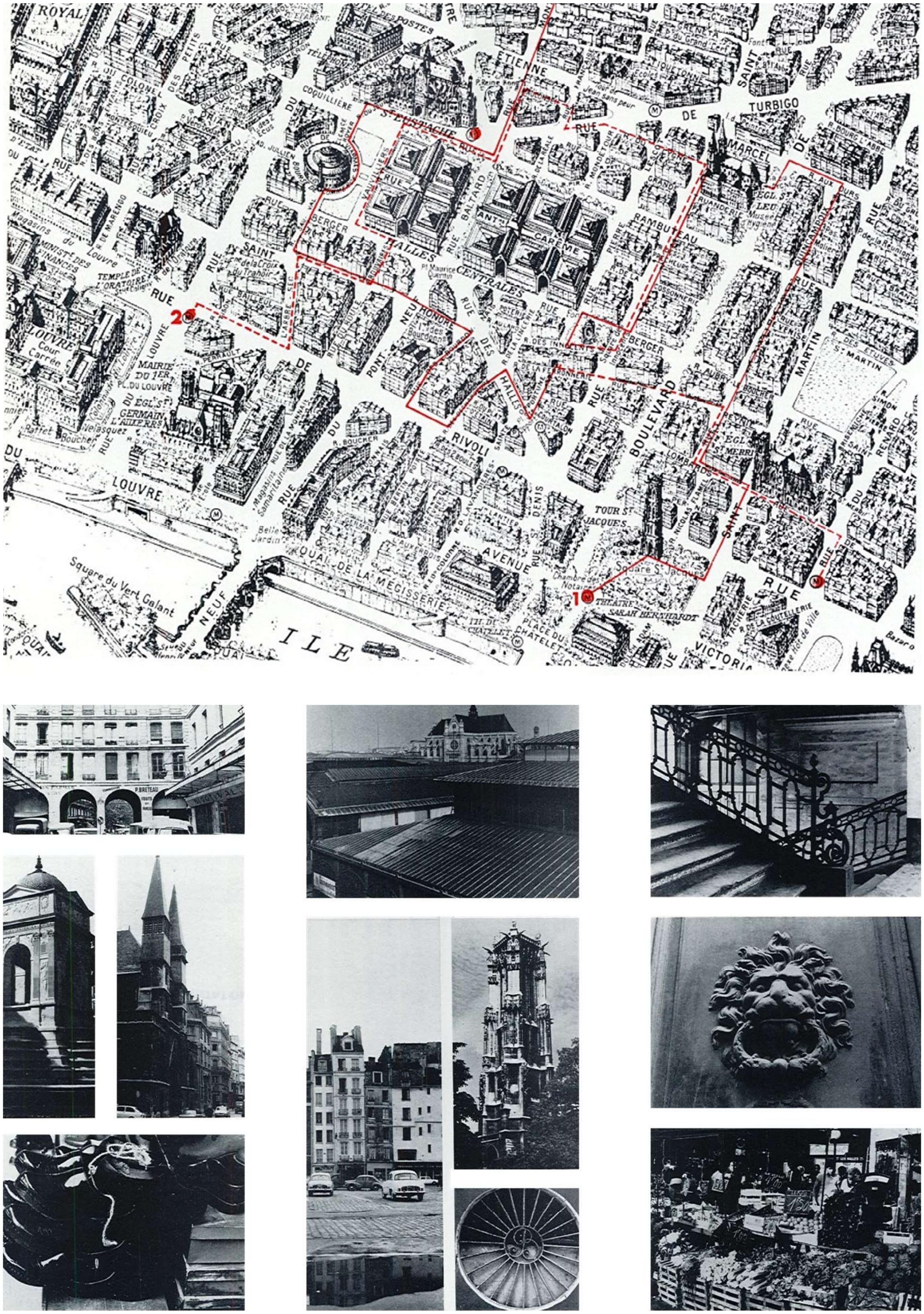
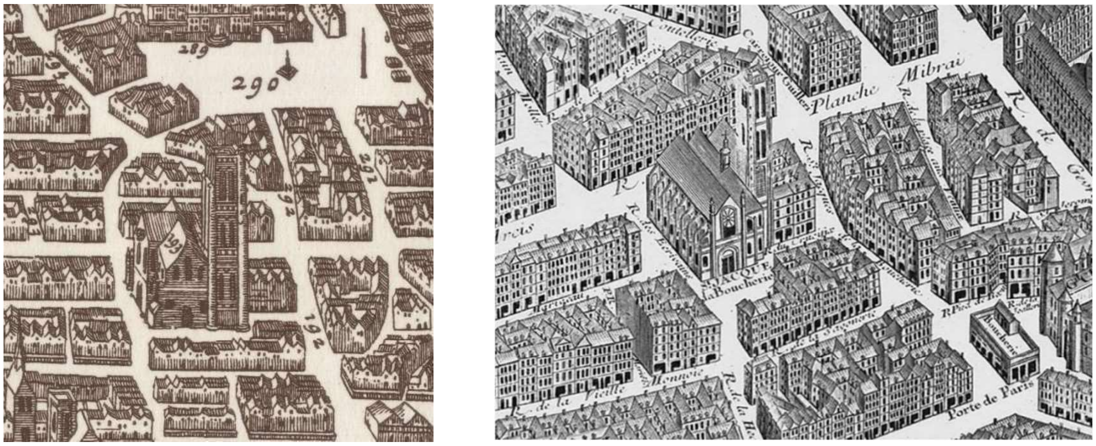


Figure 5. Maps and some buildings around Les Halles and the Tour Saint Jacques, Blondel La Rougery, 1959.

4.1 Case studies in Paris

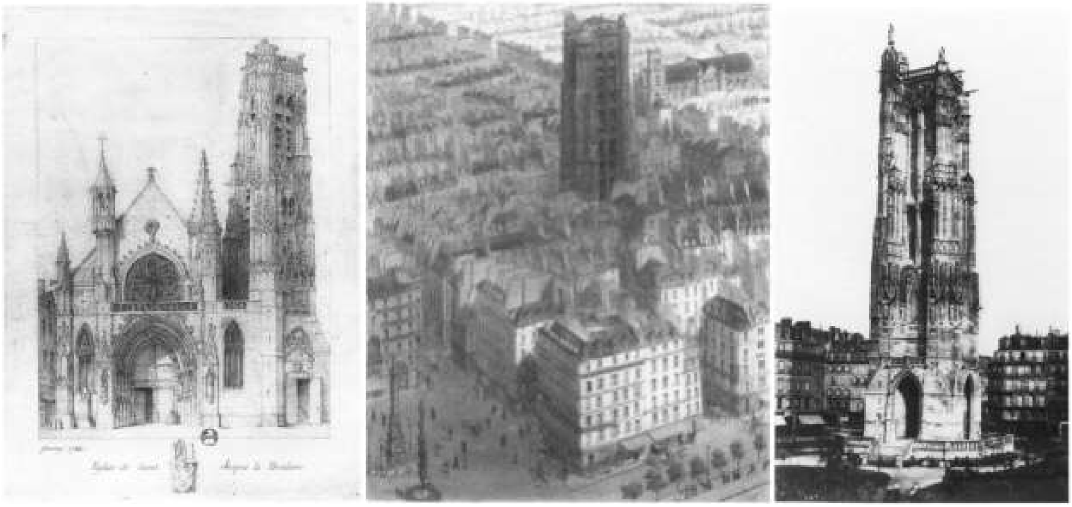
4.1.1 Tour Saint Jacques

The first selected case study is the Tour Saint Jacques (Figure 6-13) that is located in Rue Rivoli in Paris's 4th arrondissement. This bell tower is in flamboyant gothic style and it has been inscribed in the UNESCO Heritage List since 1998 for its historical importance.



(a) (b)

Figure 6. The Saint Jacques La Boucherie church. (a) Map of the city of Paris, Harold B. Lee, Scale: ca. 1:2.000, 1618, Library Maps Collection. (b) Plan of Turgot, 1734, Archives Nationales.



(a) (b) (c)

Figure 7. The Tour Saint Jacques. (a) The Saint Jacques La Boucherie church, Garnerey, 1784. (b) The tower before the isolation by restoration in the 1850s, "The Place du Chatelet and the Tour Saint-Jacques in 1848". (c) The Tour Saint Jacques on a decorative stone podium after relocation, Theodore Baldus, 1858. In O' Connell, 2001.

The Tour Saint Jacques, in fact, is the only remaining evidence of the lost Saint-Jacques-de-la-Boucherie church (Figure 6 and 7), a Carolingian chapel of the twelfth century and substantially modified thereafter, and destroyed in 1797 after civil unrest.

The tower, which over the years had been incorporated into the surroundings buildings (Figure 10 and 11) was saved from destruction because of its great architectural value. During the Haussmann transformations it was cleared by these buildings (Tamborrino, 2005) (Figure 8, 9 and 12), surrounded by a public park (Figure 13), and elevated on a decorative stone podium of 2,50 m to reach the new level of Rue Rivoli (Meurgey, 1926; O' Connell, 2001).

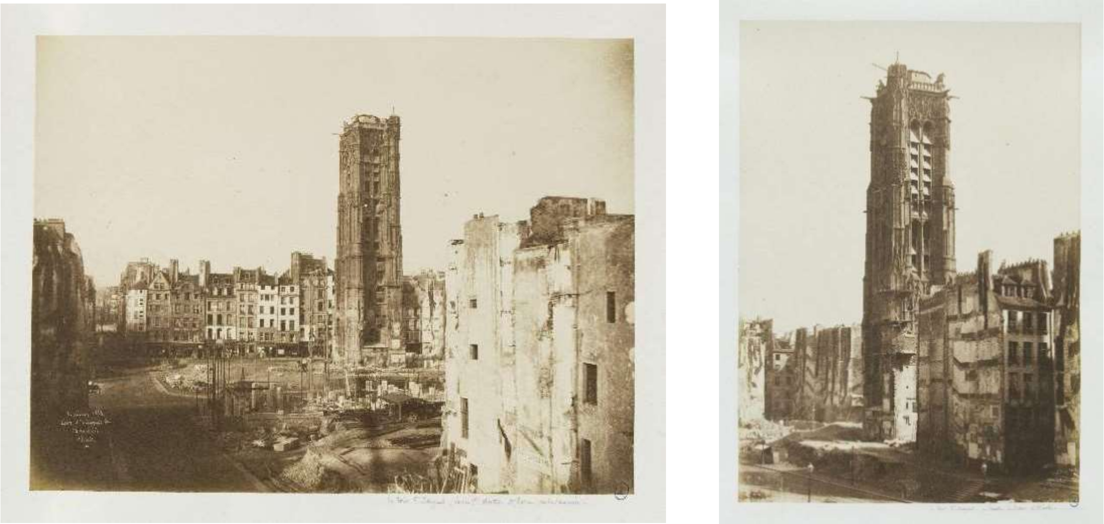


Figure 8. The Tour Saint Jacques, Henri Jean-Louis Le Secq, 1853, Musée Carnavalet.

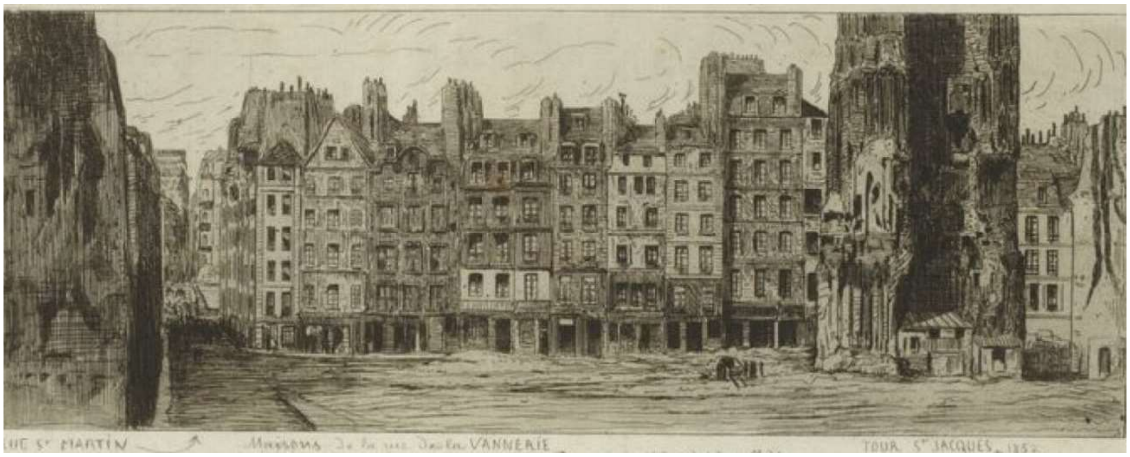


Figure 9. The Tour Saint Jacques during transformation works, view of Rue Saint-Martin and Maisons de la rue de la Vannerie, Martial, 1852, Bibliothèque nationale de France.

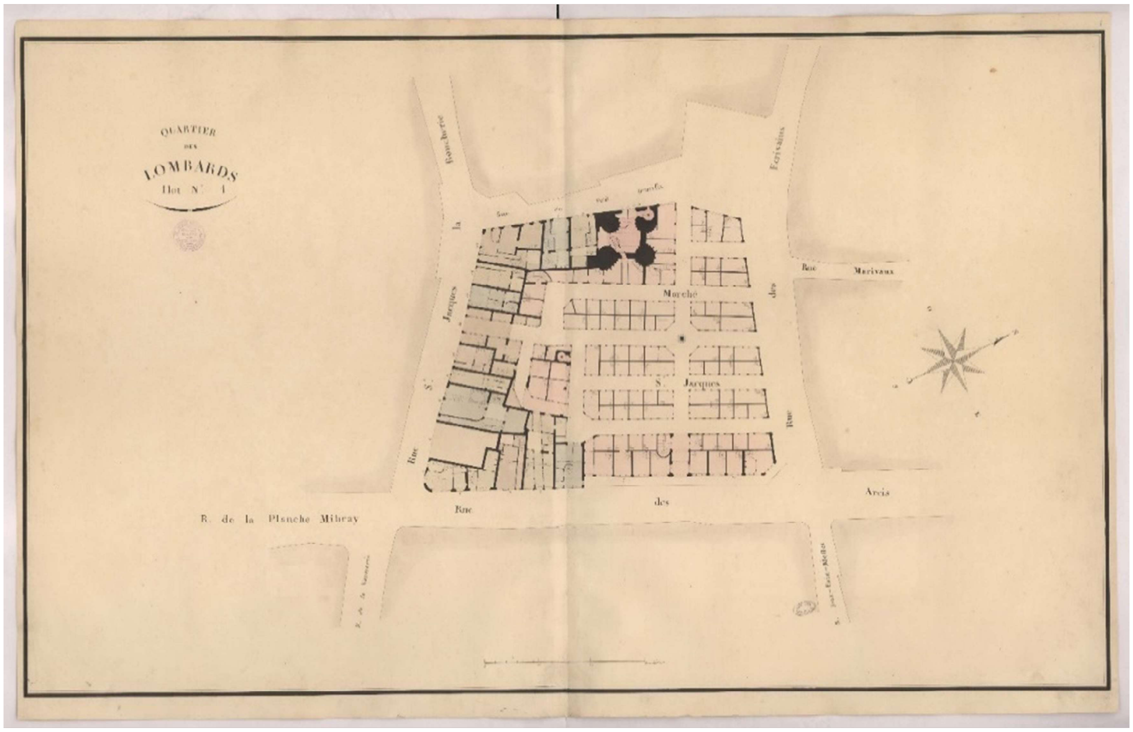


Figure 10. Plan Vasserot in which the Tour Saint Jacques is surrounded by buildings, Cadastre par îlots de Vasserot et Bellanger, 1810-1836, Archives Nationales de Paris.

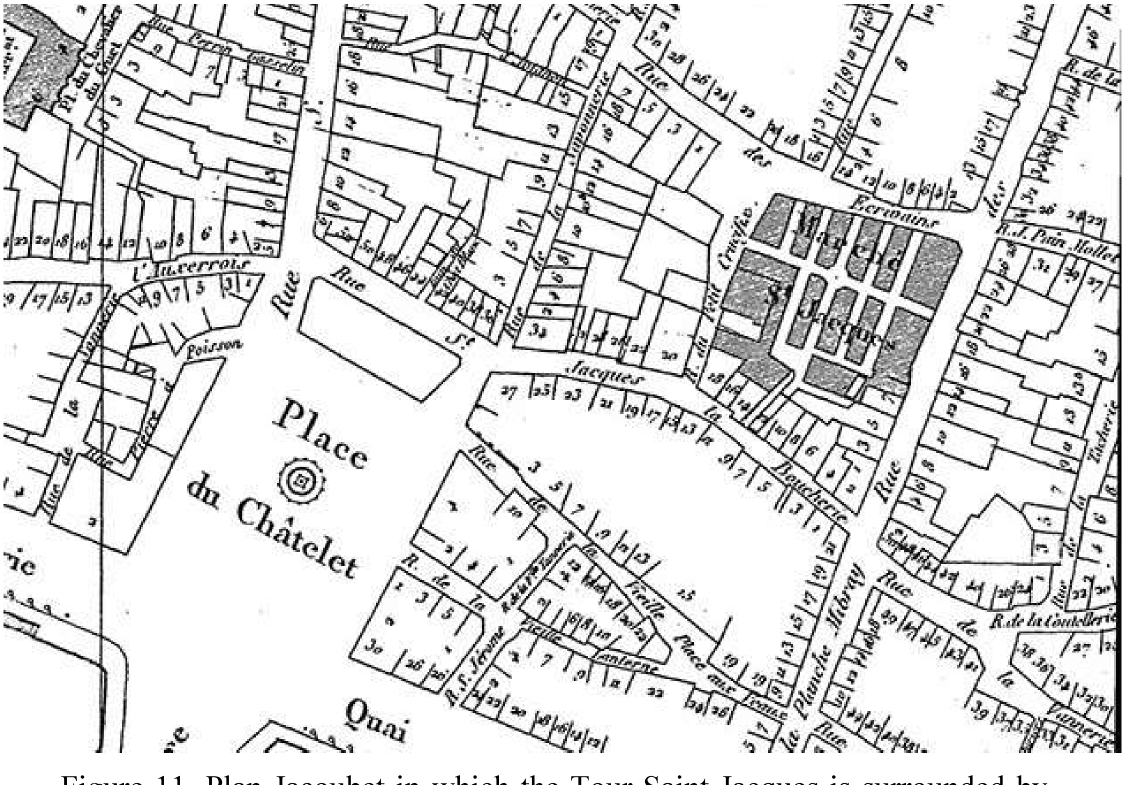


Figure 11. Plan Jacoubet in which the Tour Saint Jacques is surrounded by buildings, 1836, Bibliothèque historique de la Ville de Paris.

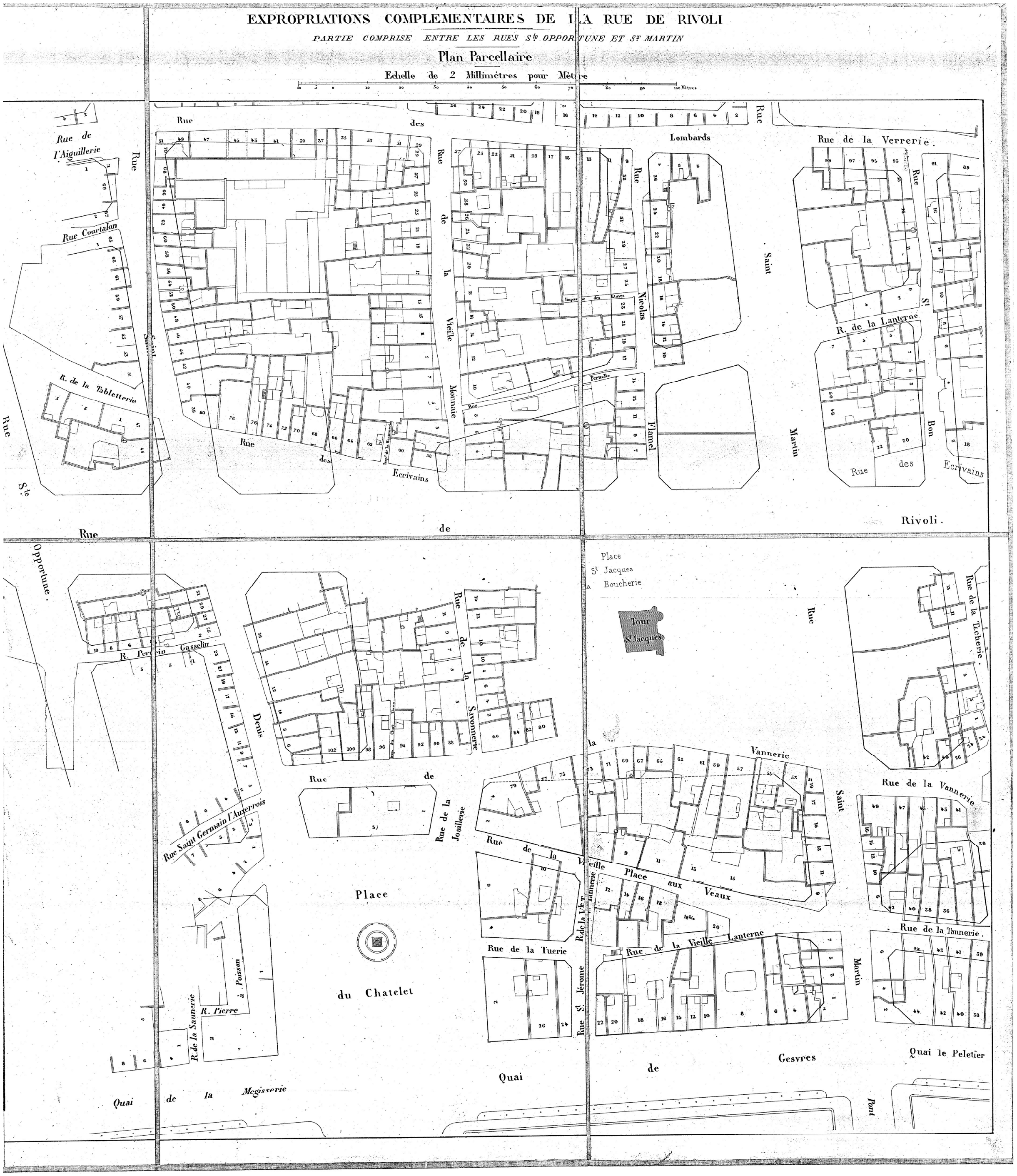


Figure 12. Plan Haussmann Paris in which the Tour Saint Jacques was moved from its position and freed by sourroundings buildings, Plan parcellaire du prolongement de la rue de Rivoli Typ. Vinchon / Gráve par Avril Frères / Lith. Lemercier, 1851.

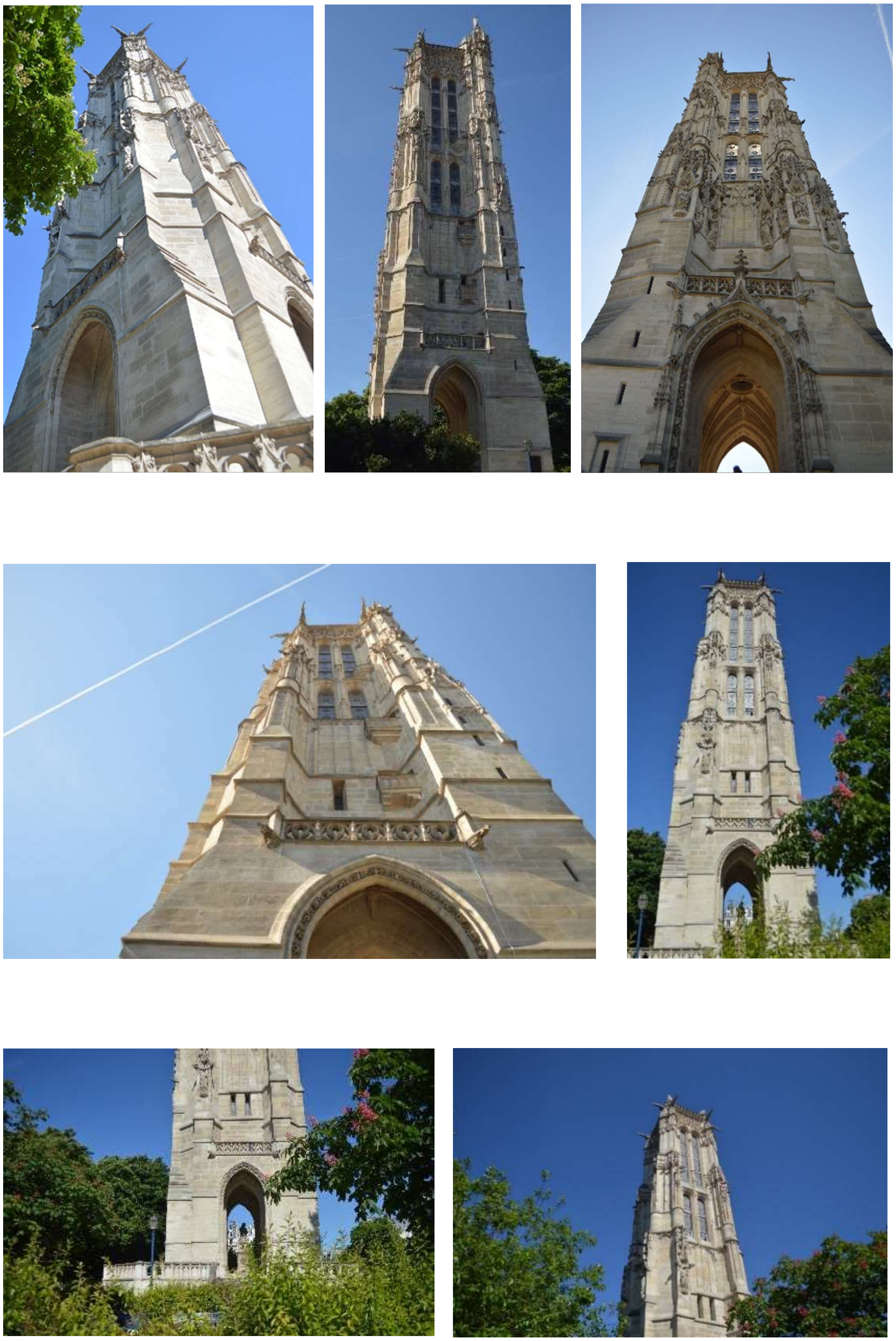
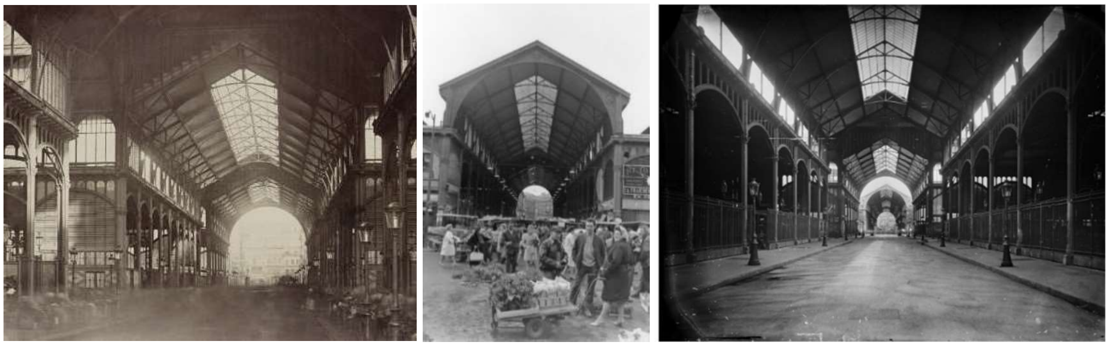


Figure 13. The Tour Saint Jacques in the actual asset, Francesca Condorelli (author), 2019.

4.1.2 Les Halles

The pavilions of the ancient market of Les Halles (Figure 14-24) were built in Paris between 1854 (Haussmann approved the project in 1953) and 1874 by Victor Baltard, one of the most important architects of the XIX century. Constituting a nerve centre in the city of Paris and object of numerous political and social debates, Les Halles have been extensively studied from the point of view of urban history due to its complexity. This complexity refers to two different aspects. The first is related to the physical place, since being a market in iron and glass, it is an indoor and outdoor space in contemporary, constituted by different layers that make the virtual reconstruction more difficult. The second aspect of its complexity regards the historical sources of different types. This aspect more interesting from the metric point of view is deeply investigated in this work, focusing on the combination and the metrical exploitation of the historical material thanks to the use of photogrammetry and DL.



(a) (b) (c)

Figure 14. Les Halles. (a) Charles Marville, 1855, Musée Carnavalet / Roger-Viollet. (b) and (c) 1943, Musée d'Orsay.

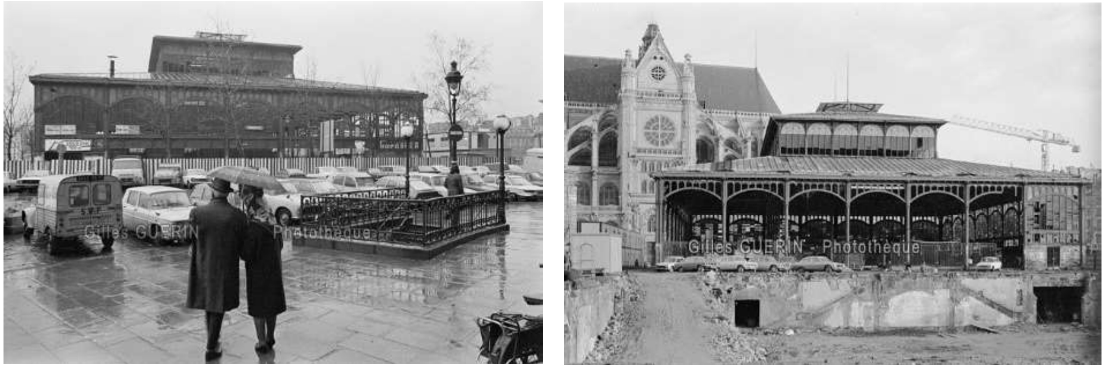


Figure 15. Les Halles, Gilles Guérin, 1973, (http://www.gillesguerin.com/).

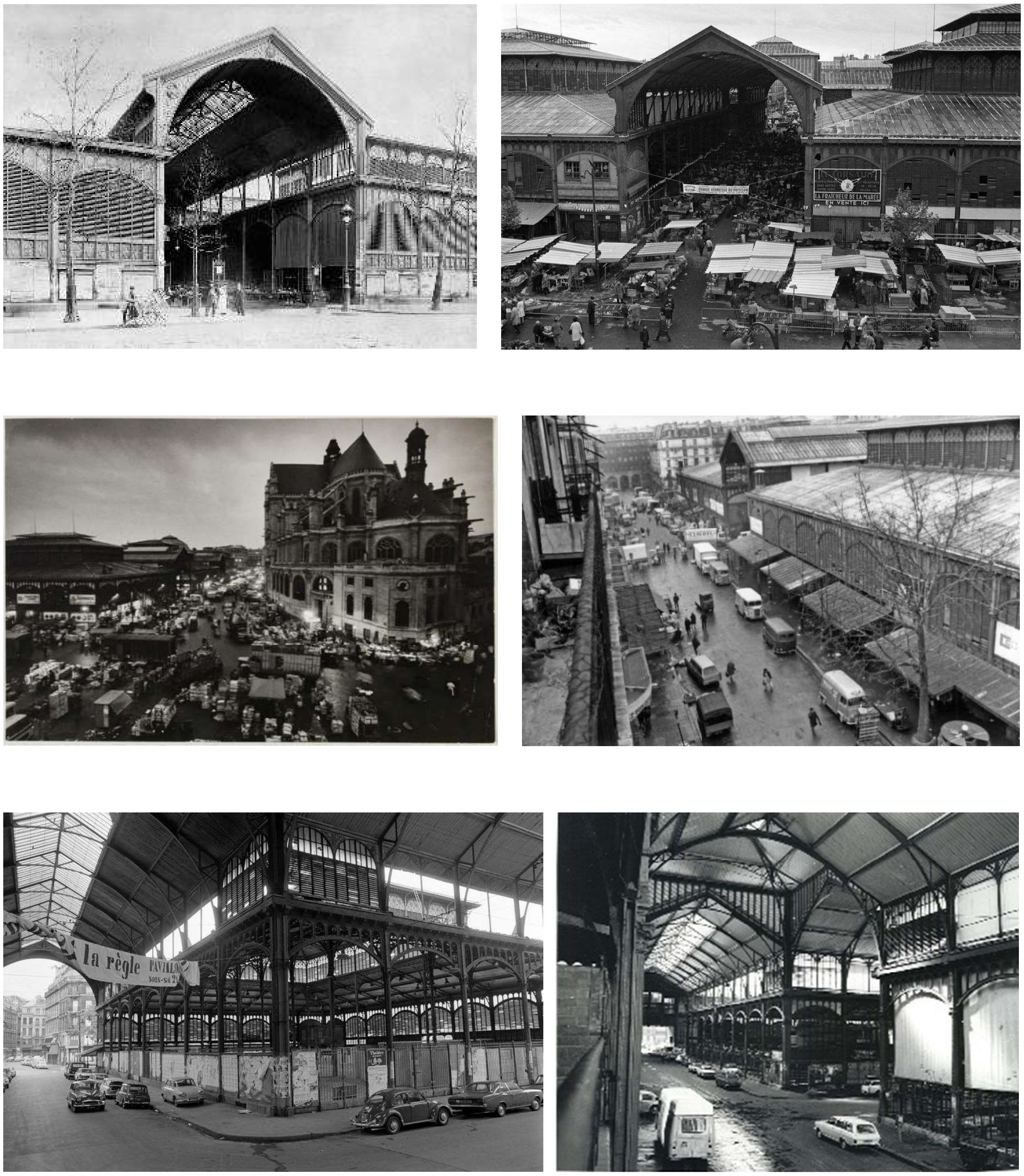


Figure 16. Les Halles, 1943, Musée d’Orsay and Musée Carnavalet / Roger-Viollet.

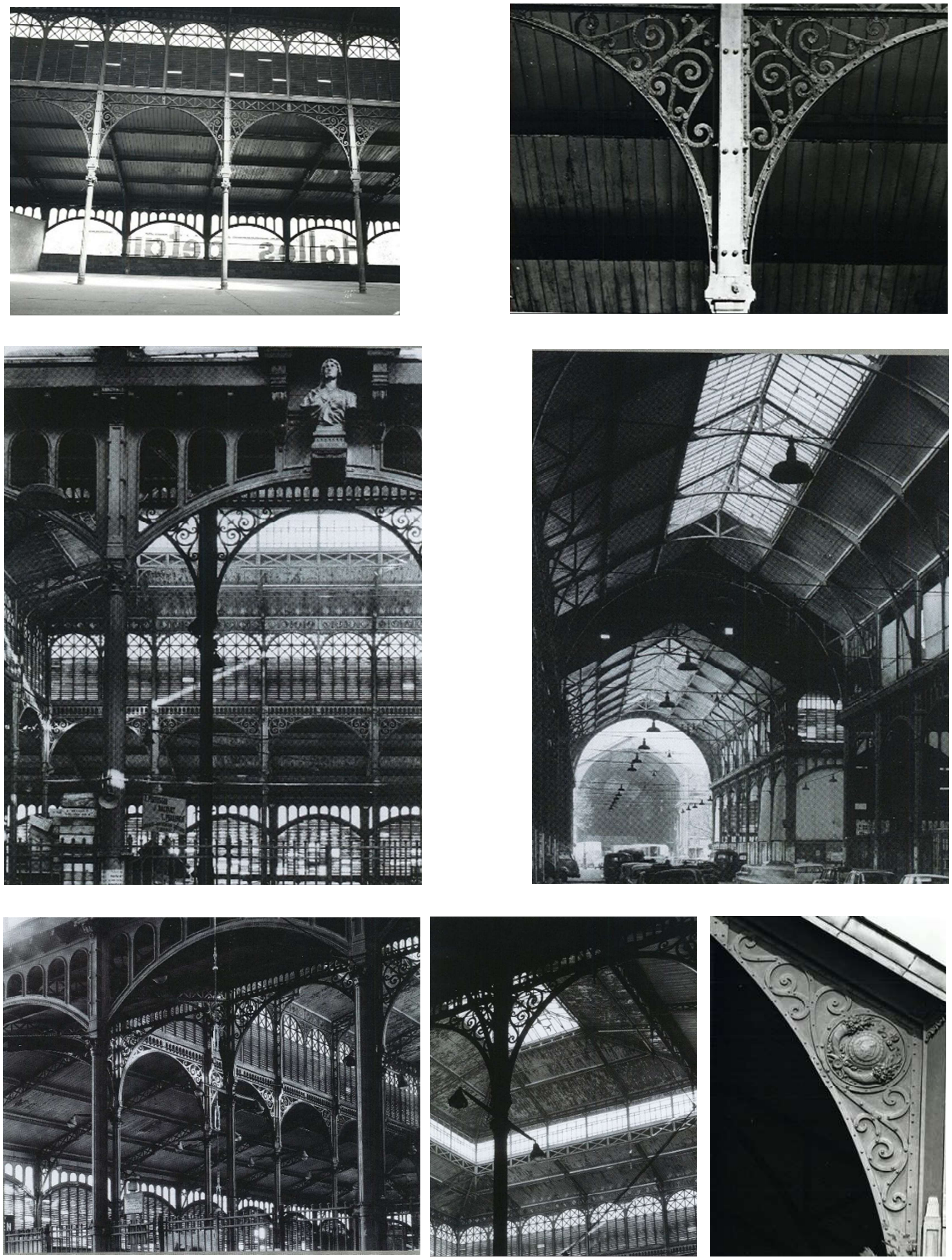


Figure 17. Details of Les Halles (in Thomine-Berrada, 2012).

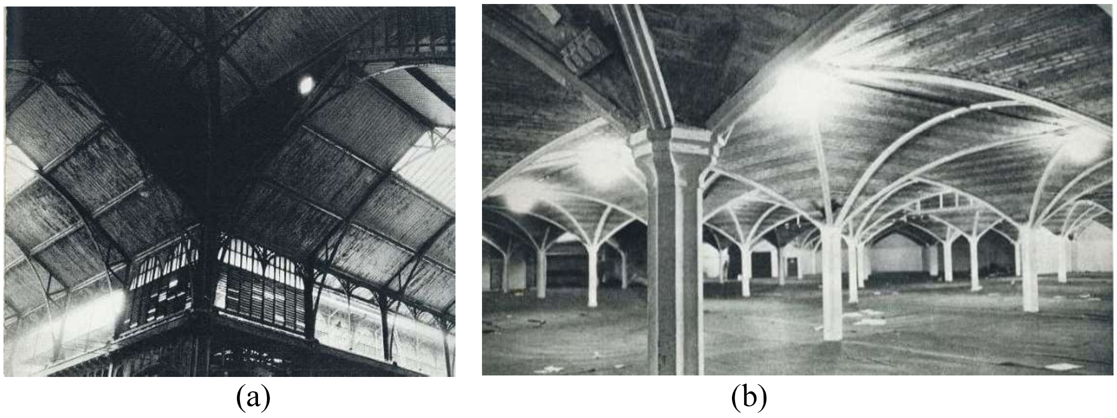


Figure 18. Les Halles. (a) Detail of the roof. (b) The cave. (Medecine de France N° 226, 1971).

Les Halles by Baltard were demolished in 1971 (Figure 19 and 20) and now it is no longer possible to observe them in situ. After the transformation of the area, a new building, Forum des Halles, was designed in 1979 and then restructured in 2002. After a century of activity, the need of a more modern market, in line with the challenges of supplying society with a rapidly growing population, arose. Congestion, dilapidated housing in the neighbourhood, the monopoly of the streets, traffic, hygiene and demographic pressure have been causing daily problems in the Les Halles neighbourhood for many years. On 6 January 1959 the government decided to move them to a more appropriate space outside the city. Ten years later and after a long implementation, the transfer began on 3 March 1969 and ended in January 1973 (Archive of Paris).

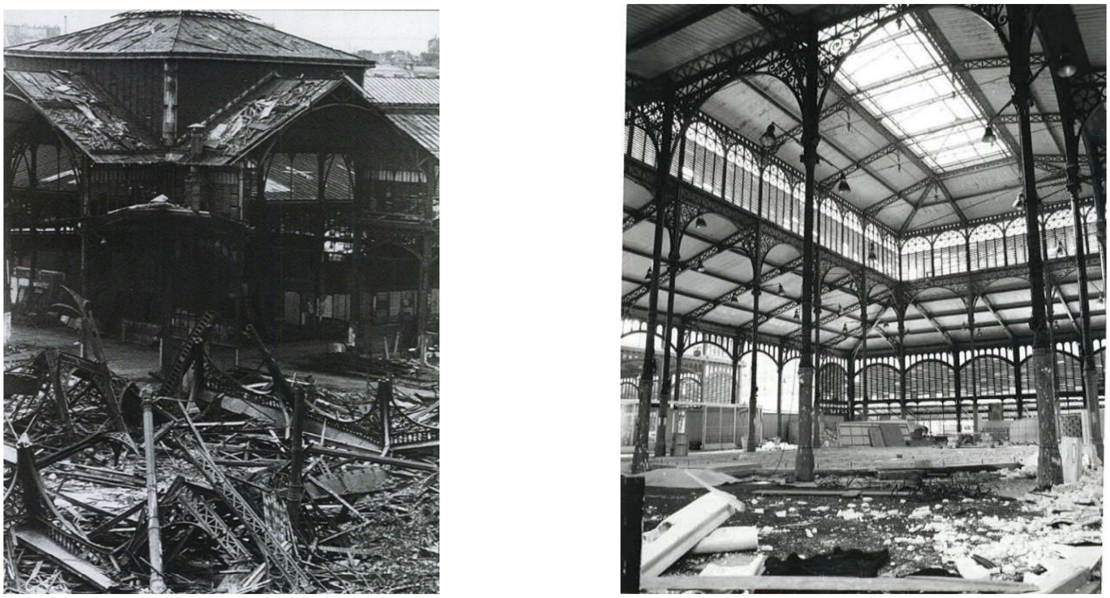


Figure 19. The destruction of Les Halles (in Thomine-Berrada, 2012).

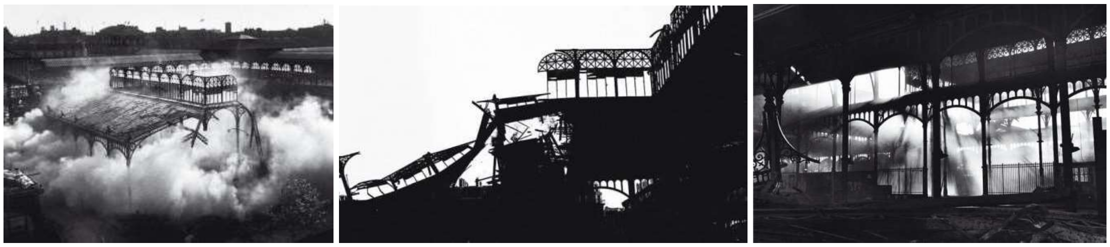
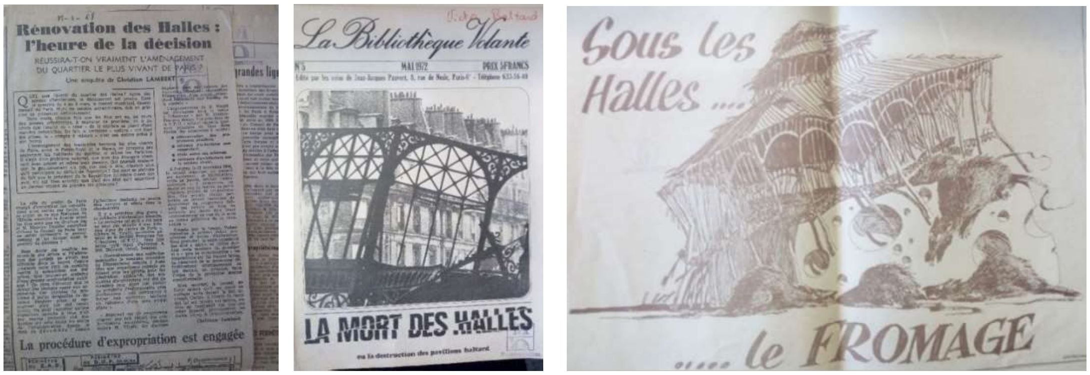


Figure 20. The destruction of Les Halles, Jean-Claude Gautrand, 1977, Galerie W.

This decision provoked significant reactions from Parisians who signed petitions against the destruction of the pavilions (Figure 21) and the expropriations caused by the renovation of the neighbourhood. It is the end of an activity that has become legendary in the landscape of Paris. Les Halles soon became an admired monument, a must for any foreign tourist passing through Paris. Paris' guides contained a chapter on Les Halles, whose appearance and activities they describe. Passengers passing through do not fail to mention their visit to Les Halles. Several articles were dedicated to the pavilions during and after their construction and popular literature found great phrases to qualify Les Halles as "true cathedrals of cast iron and glass, agile and light, in their unchanging solidity, luminous and airy".



Rénovation des Halles : l'heure de la décision

RÉUSSIRA-T-ON VRAIMENT L'AMÉNAGEMENT DU QUARTIER LE PLUS VIVANT DE PARIS ?

Une enquête de Christian LAMBERT

La procédure d'expropriation est engagée

La Bibliothèque Volante

N°5 MAI 1972 PRIX 5 FRANCS

LA MORT DES HALLES

ou la destruction des pavillons baltard

Sous les Halles ...
... le FROMAGE

Figure 21. Example of newspaper against the destruction of Les Halles, 1972, Musée d'Orsay.

The literary commemorations of Les Halles are numerous, but dominated by the monument that Emile Zola consecrated to them. The genesis of the novel "Le ventre de Paris" dates back to 1869. Zola is fascinated by the swarm of life that animates those pavilions. Les Halles are the protagonists of this fragrant symphony, of this violent and serious painting of humanity that is beastly and complex.

They are among the latest demolitions carried out in the '800 in Paris as they have generated a lot of discontents. For example, Musée d'Orsay was saved from destruction after this event. Cultural institutions wondered how to preserve this architectural heritage. Many reports in the written press or on television want to inscribe images of Les Halles' life and activities in the collective memory and insist on the "end of a world". The photographs and videos taken during that period contribute to this memory. The whole of France followed the immense site of destruction of the pavilions in the media of the time.

Finally in 1972 it was decided to restore Pavilion No. 8 (Figure 22), which has been saved and reassembled with some modifications in two different location, one is in Nogent-sur-Marne, Île-de-France, and the other is in Yokohama, Japan.

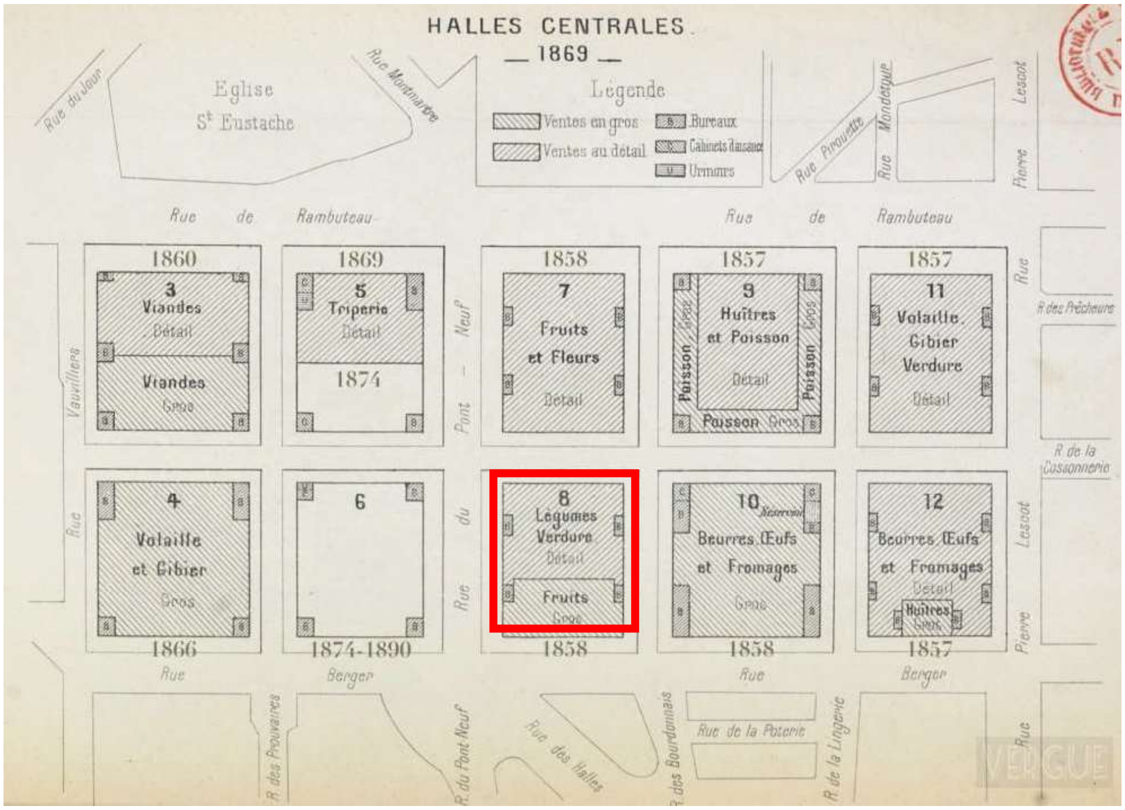


Figure 22. Plan of Les Halles, 1867, Bibliothèque des Archives de Paris. The Pavilion No. 8 highlighted in red was moved in Nogent-sur-Marne and in Japan.

The main part of the pavilion was moved in Nogent-sur-Marne (Figure 23) and used for exhibitions, salons, concerts, etc. Various modifications were made to adapt it to the new use: installation of blue mosaic tiles that take up the motifs of the old brick panels, closing of the facades with Parol glass, installation of the heating system, thermal insulation of the rebuilt roof, arrangement of mobile scenographers. The dismantling started in 1872 and ended 3 months later. Removal in 1976 is inaugurated in 1977.

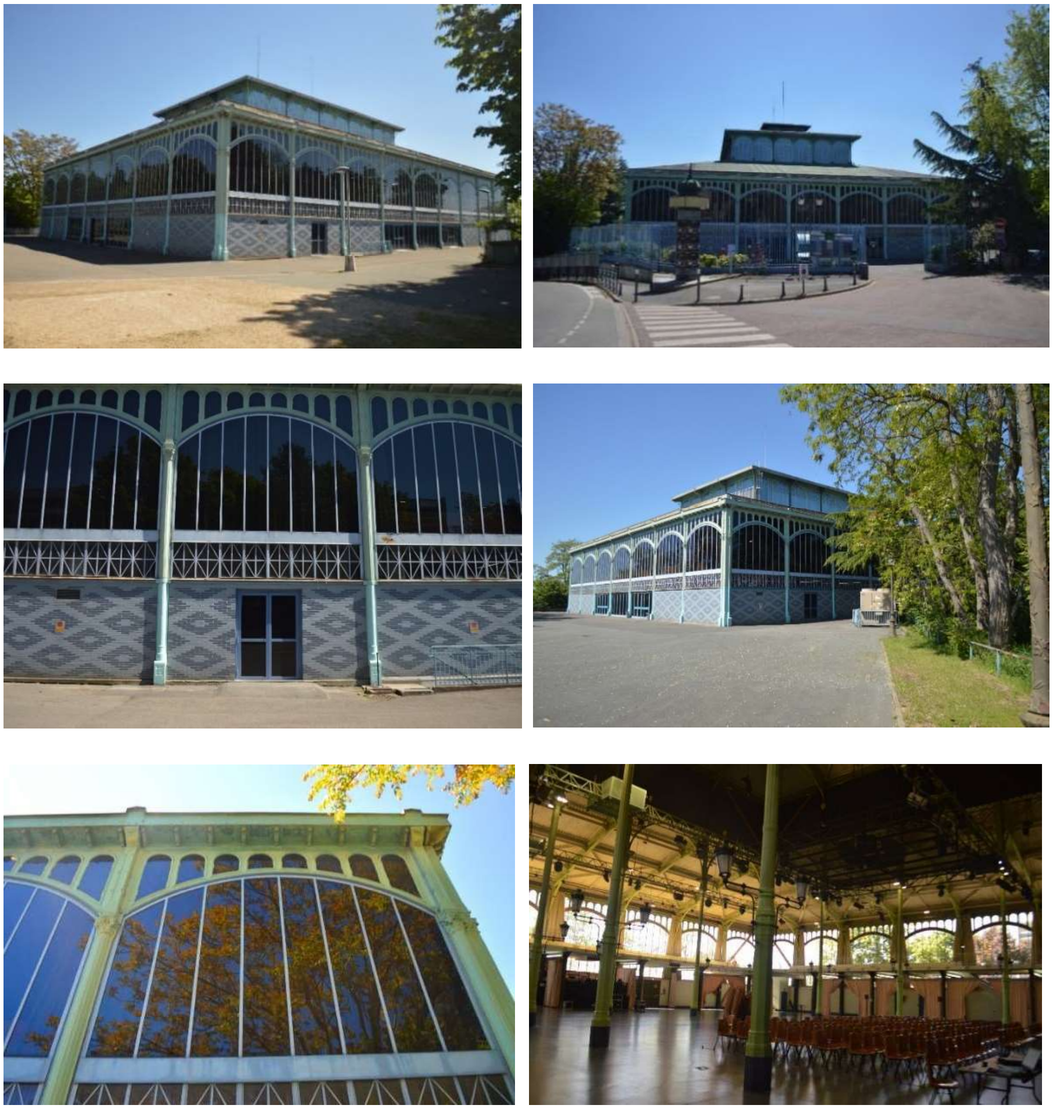


Figure 23. Pavilion n°8 in Nogent-sur-Marne, Francesca Condorelli (author), 2019.

The cave of Pavilion No. 8 (Figure 24) which was to support the upper structure of the basement portion, is composed of all such as pillars, arches and cast iron beams was donated as a courtesy of the Commune of Paris to the city of Yokohama and located on the Yamate hill. The French Consulate was built on the same site in 1894 and was in use until the 1940s. The site was purchased from France by the City of Yokohama, and it built the present park, naming it France-Yama. The remains of the consulate, which was destroyed by fire, still stand halfway up the hill. The ruins of the Consulate and a replica of the windmill.

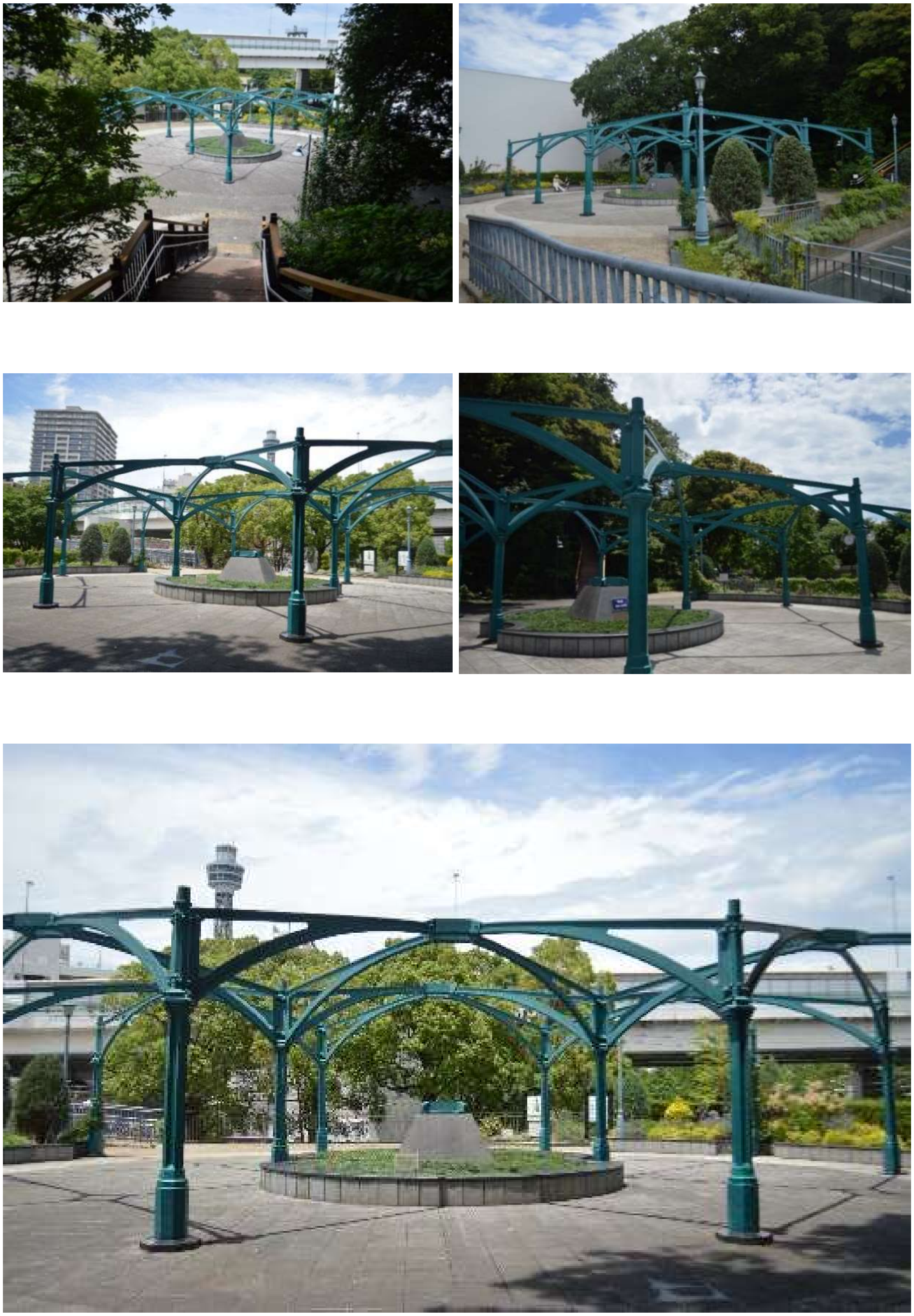


Figure 24. Pavilion No. 8 in Yokohama (Japan), Francesca Condorelli (author), 2019.

4.1.3 Recovering the material on the case studies

Numerous historical material regarding the monuments survived over time. For this reason, different historical archives were deeply consulted in Paris in order to collect documents concerning the two case studies. Historical photographs, drawings, images and design projects have been collected from the following archives: the Musée d'Orsay archive, the Archives de Paris, the Bibliothèque Nationale de France, Archives Nationales de France.

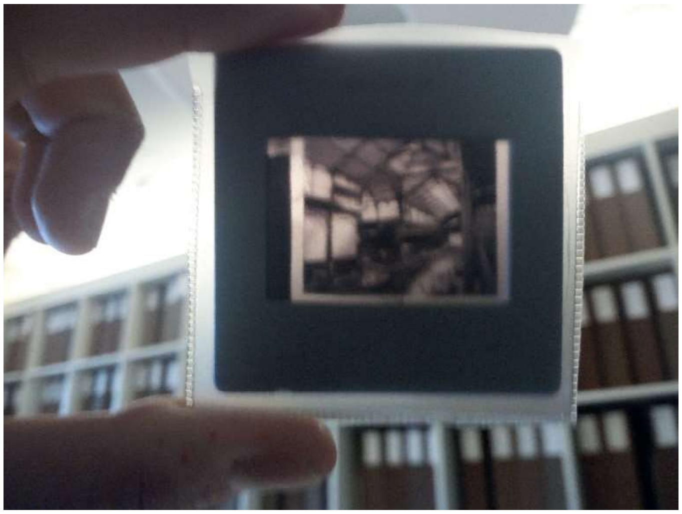


Figure 25. A diapositive of Les Halles at Musée d'Orsay, Francesca Condorelli (author), 2019.

In particular, for the Tour Saint Jacques, together with maps and photographic reportage of its transformation during the time, a collection of drawings by the architect Gabriel Davioud of the surroundings of the tower was available thanks to the participation in the "Summer School: Cities Cultural Heritage and Digital Humanities" organized by prof. Tamborrino. In order to preserve at least the memory of the medieval houses that were to be demolished, the Haussmann administration had taken care to preserve the graphic memory of the expropriated districts documenting with drawings made between 1840 and 1860 of the buildings in the area (Tamborrino, 2005). The collection of drawings consists of elevations drawn in pencil on a scale of 1:100. They show many details of the buildings surveyed and offer an image of Paris lost in the middle of the 19th century. Moreover a 3D model, obtained from a recent photogrammetric survey of the existing Tour Saint Jacques using UAV carried out by Iconem in 2015 was kindly made available from the company, and a 3D model of the existing pavilion of Les Halles by the archive of the Commune de Nogent-sur-Marne. This material was used for the metric quality assessment, as better explained in the next section 4.2.

For what concerns Les Halles, plans, sections, elevations and details have been published several times by the architect Victor Baltard himself in an important monograph. All this allows the offering of a fairly faithful description of the way in which Les Halles were built.

The “Monographie des Halles centrales de Paris” by Baltard, Archives de Paris, ATLAS 97, is the most complete and accurate publication in existence on Les Halles; it provides documentation of the 12 pavilions designed (the last two were only built in 1936 and not in accordance with the project).

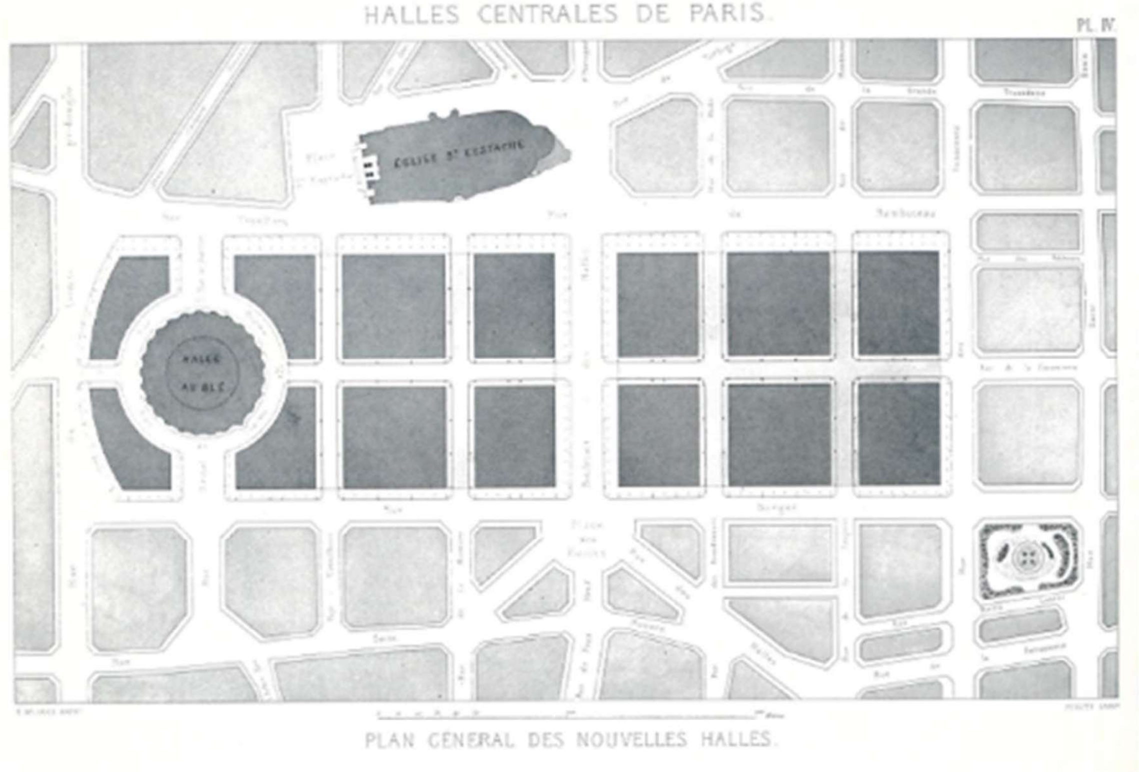

HALLES CENTRALES DE PARIS.
PL. IV.
EGLISE St EUSTACHE
HALLE AU BLÉ
PLAN GENERAL DES NOUVELLES HALLES.

Figure 26. Plan of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris.

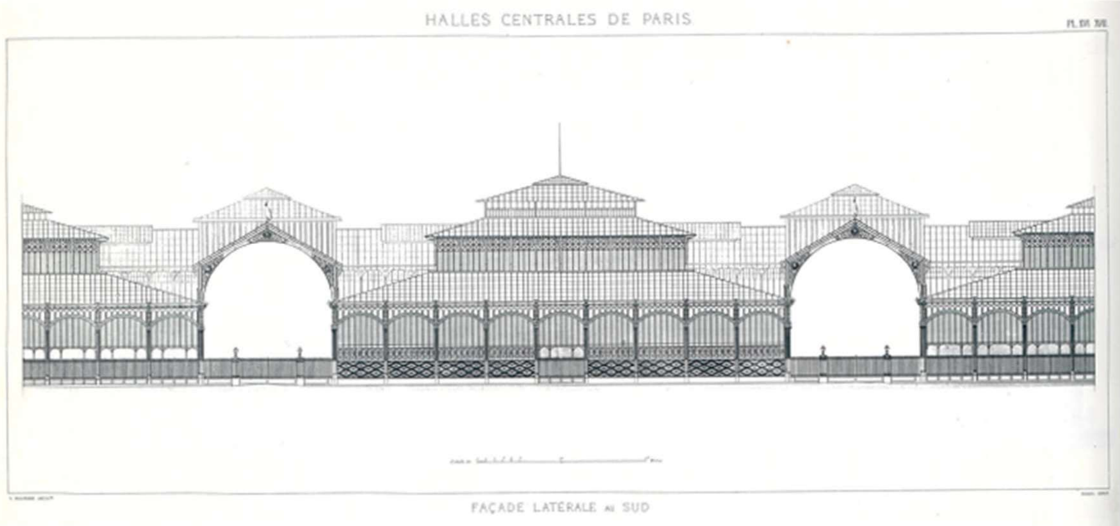


Figure 27. Façade of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris.

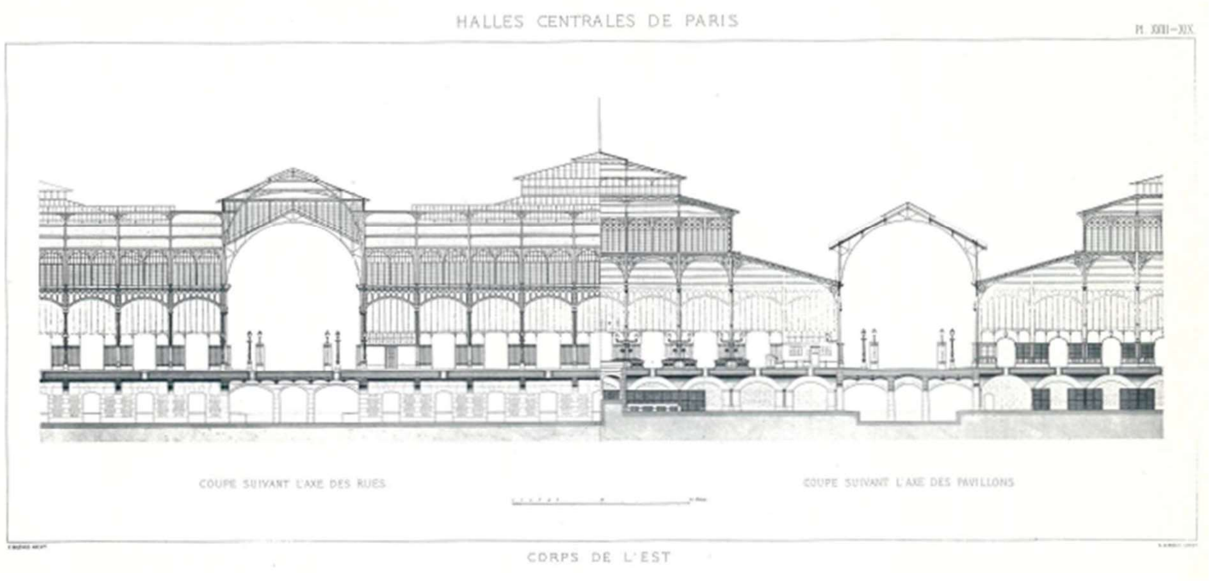

HALLES CENTRALES DE PARIS
Pl. XVIII–XIX
COUPE SUIVANT L'AXE DES RUES
COUPE SUIVANT L'AXE DES PAVILLONS
CORPS DE L'EST

Figure 28. Section of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris.

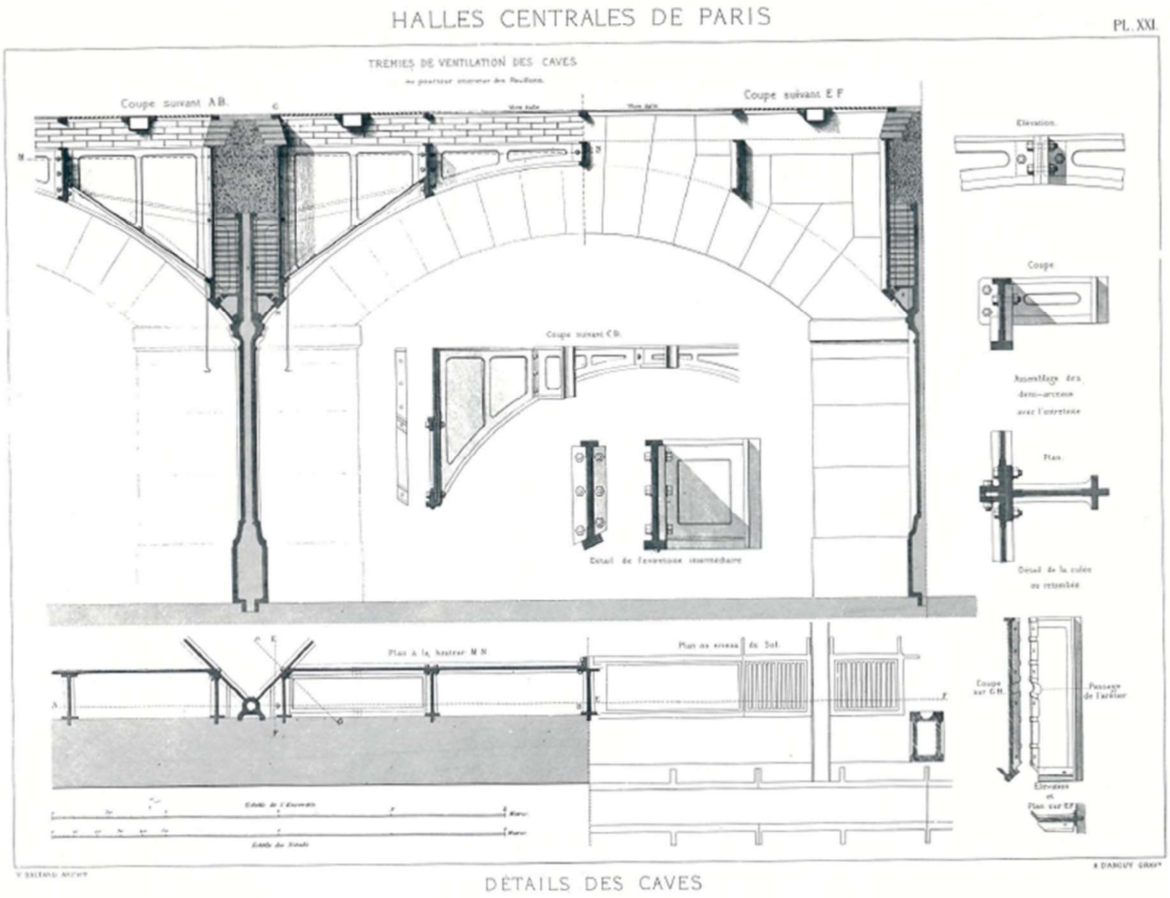


Figure 29. Details of the cave of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris.

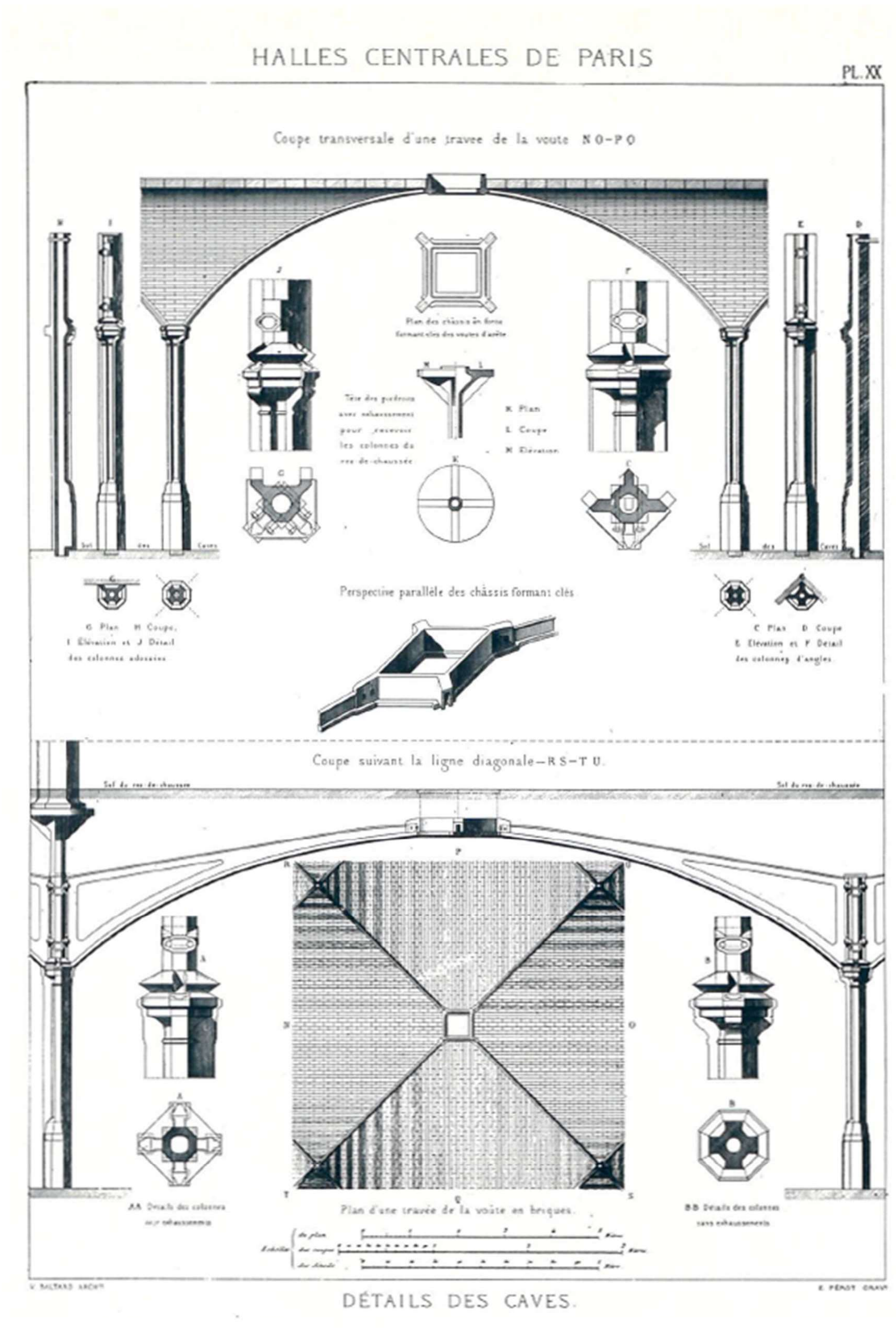

HALLES CENTRALES DE PARIS
PL. XX
Coupe transversale d'une travee de la voute NO-PO
Plan des châssis en fonte formant clés des voutes d'arête
Tête des piédroits avec rehaussement pour recevoir les colonnes du rez-de-chaussée
K Plan
L Coupe
M Élévation
Sol des Caves
Sol des Caves
Perspective parallèle des châssis formant clés
G Plan H Coupe, I Élévation et J Détail des colonnes adossées
C Plan D Coupe E Élévation et F Détail des colonnes d'angles
Coupe suivant la ligne diagonale—RS—TU.
Sol du rez-de-chaussée
Sol du rez-de-chaussée
P
R
U
N
O
T
Q
S
AA Détails des colonnes avec rehaussements
Plan d'une travée de la voûte en briques.
BB Détails des colonnes sans rehaussements
Echelles du plan des coupes des détails
V. BALTARD ARCHTE
E. PÉROT GRAVR
DÉTAILS DES CAVES.

Figure 30. Details of the cave of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris.

In Lemoine, 1980 is also reported a description of the pavilions' structure containing metric information: "The excavation of the foundations was 6.70 m to allow the passage of the railways, then it descended to 7.70 m at the level of the turntables. The construction system of the cave was simple and at the same time intelligent. A number of cast iron support points at a distance of 6 m from each other and arranged alternately, supported a network of sharp-edged vaults whose ribs were also made of cast iron. The poles had an octagonal section and an average diameter of 265 mm, increased to 440 mm for those that had to support the roofs. Originally the vault reinforcement was made only of cast iron ribs. Subsequently, another reinforcement tangent to the curve of the dyspluvium was added. In the keystone, a cast iron frame, which contained a grid or glass plate, joined the two half trusses. This system, which actually consisted of 4 cranes mounted rigidly with their hooks, was thrust free. The vaults were made of a single thickness of Burgundy bricks put together in copy, in bands of two colours. The sides of the vaults were filled with cement. Under the covered streets the thickness is 44 cm, corresponding to rows of bricks. At the intersection of the streets, which form a square of 15 m side, 1 m high sheet metal beams support a series of brick vaults. Brick was widely used in the external parts of the pavilions. The fence is solved with brick partitions 11 cm thick and 2.60 m high, which rest on a base of red Vosges stoneware. This arrangement protects the inside of the pavilions from draughts. These brick partitions are naturally supported by cast iron columns that support the iron structure of the pavilions. The combination of the metal with brick, later typical of covered markets and industrial uses, was something really new at that time. The frame of the pavilions rested on master columns and on a row of central columns at a distance of 4 m from each other that followed the plan of the shops. The zinc roof covering rested on a double layer of planks separated by a cavity to improve thermal insulation. The structure of the low sides of the pavilions was made up of a series of iron truss beams, with I-section, resting on cast iron columns and iron pedestals with a roughly octagonal section that extends the central columns. Some cornerstones, adorned with rosettes, brace the device. These same pedestals were connected to each other by arched corbels of a circumference sufficiently large and rigorously joined to the mountain arch and to the crossbeam in order to be absolutely opposed to the closing and opening of the corner. On this crossbar, placed at a height of 12.50 m, stands the first level of the lamppost. A flat iron terrace around each pavilion allowed the arches to be reached. The first level of the lamppost was surmounted by composite trusses: a series of struts rest on an inverted truss beam on the four sides of the pavilion subtended by a round iron tie-rod. The same device was repeated for the second level of the lamppost, on 18 m instead of 30. A glass skylight chorus to everything, culminating about 25 m above the ground. The small pavilions did not have a two-level streetlight but a simple glass skylight" (Lemoine, 1980).

The Tour Saint Jacques and the parts of the existing Pavilion were photographed during a visit both in Paris and in Yokohama.

Finally, both architectures were deeply studied and documented with both photographic and video material, because existing after the invention of the cinema and they appear in historical film footage. Moreover, the availability of historic documents such as plans, project and drawings with metric information, have helped the metric evaluation process.

In this way, it is possible to compare the different results obtained in the implementation of the workflow with the two case studies and therefore, they represent good case studies to test the proposed algorithm.

Historical video archives consulted

A lot of video materials, both documentary and fictional, which were set in the market and near the tower, were examined in several video archives in Paris and historical film footage from the 1910s until the 1970s were collected. The video archives consulted in Paris during a research period abroad are reported following.

- Lobster Films (last access October 2020, https://www.lobsterfilms.com/fr/), created in 1985, is a company whose goal is to restore films and share their discovery of the cinema from the early days until today. The cinematheque has a vast catalogue also containing rare films.

- CNC (last access October 2020, http://www.cnc-aff.fr/) collects, inventories, saves, restores and indexes the films it preserves through voluntary deposits, donations, acquisitions and legal deposit. The online documentary database offers a selection of films, regularly updated, from among the 110,000 films stored. Of these, 7,410 have been digitized and made available to researchers and professionals at CNC consultation stations in Paris and Bois d'Arcy. The collections are open to many national and international partners and accessible to the public under certain conditions while respecting the rights to the works.

- Institut national de l'audiovisuel – INA (ina.fr, last access October 2020) is a public commercial entity in France responsible for archiving all radio and audiovisual broadcasts in the country with particular focus on conservation, restoring, accessibility on the internet, enhancing archives for scientific, educational and cultural purposes of the French audiovisual heritage.

- Les Documents Cinématographiques (https://www.lesdocs.com/, last access October 2020) is a society created in 1930. The cinematographic archive whose origin dates back to 1889 it now includes a catalogue of historical films, documentaries and fiction, intended for the general public and also for audiovisual professionals, particularly in the field of television rights and sales.

- Forum des Images (https://www.forumdesimages.fr/, last access October 2020) founded in 1988 is a cultural institution of the city of Paris dedicated to cinema and audiovisual. Its archives include almost 8,000 films with Paris as subject or setting, as well as several hundred other documents from various rare and difficult to access collections.

- Gaumont Pathé Archives (https://gparchives.com/, last access October 2020) was set up after the catalogues of Cinémathèque Gaumont and Pathé Archives were combined in 2004. This venture is now the leading French image bank for black and white and colour images illustrating the history of the 20th and 21st centuries. The archive contains nearly 12,000 hours of footage and numerous documentaries.

All the historical film footage collected and used in this research were already digitalized by the archives. According to the archive expert technicians, the conversion procedure into digital format of the film did not conditionate, in an appreciable way, his quality and characteristics, since the process was carried out with high quality instruments and well-specialized method. Metadata of the films not always are present in the cataloguing database, for this reason, the information necessary for the photogrammetric processing were obtained according to the following considerations:

- When the video camera and the film used to shoot the video are known, the image format and the focal length can be found out. The image format is, in fact, related to the type of film used (for example: the film of 35 mm format has a height of 35 mm and the dimensions of the frame are 18 mm x 24 mm, as shown in Figure 31). The focal length is related to the camera used and generally has a fixed-focus lens.
- When the metadata on the camera is unknown, the missing information was found out searching the camera available in the year and in the place in which the video was taken (considering that at the beginning of the ‘900 only few cameras were invented).

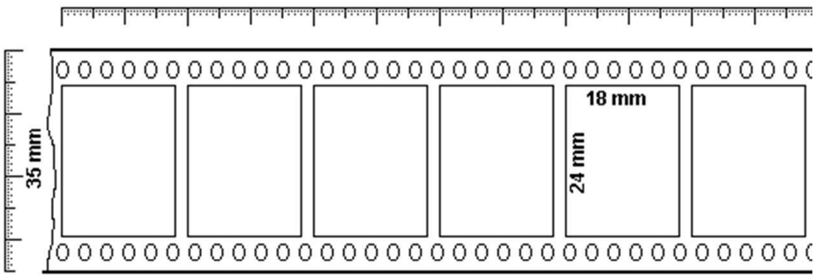


Figure 31. dimensions of the 35 mm film format.

4.2 Testing the entire workflow: AI + photogrammetry

The case study selected were used to experiment with the entire workflow prosed in this dissertation and reported in Figure 32. In particular, the existing monument of the Tour Saint Jacques was adopted to test the implementation of two different types of Neural Networks with different training scenarios in order to determine the best performance between the two. This required a large number of images, which were used together with historical and contemporary images of the monument for both the training and validation phases. Once the experimentation of the two Neural Networks had been carried out and the relative results analysed, the same methodology was tested by applying the results obtained in this first part to a real case, such as the recognition of a monument lost in a film, Les Halles, which no longer exists. For this reason, the two case studies of the tower and the pavilions were analysed in parallel, in order to gain a deeper insight into the minimum number of images required to form a network.

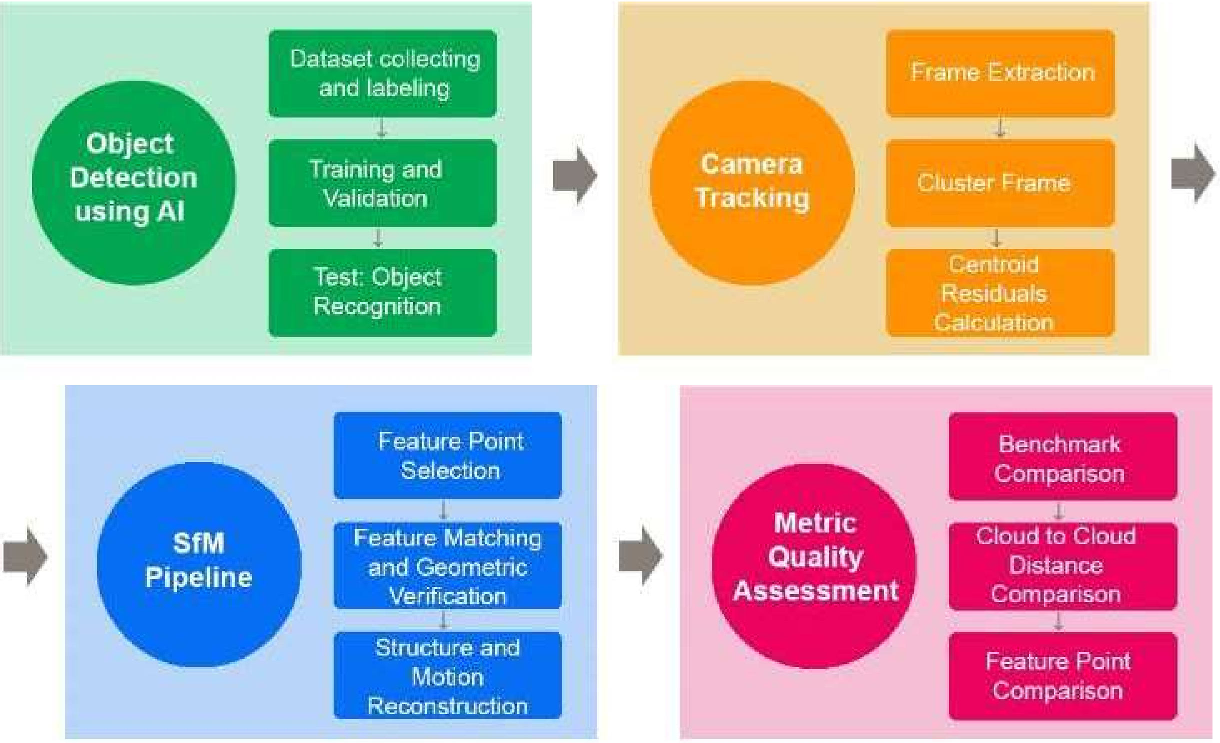

Object Detection using AI
Dataset collecting and labeling
Training and Validation
Test: Object Recognition
Camera Tracking
Frame Extraction
Cluster Frame
Centroid Residuals Calculation
SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction
Metric Quality Assessment
Benchmark Comparison
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 32. The workflow proposed in this dissertation.

4.2.1 Dataset

The implementation of the Neural Networks needs a specific dataset from which to learn to recognize the searched object. For the cases of the Tour Saint Jacques and Les Halles, even if a lot of data were available, no suitable datasets for the application of Deep Learning (DL) existed, and a new one was specifically created.

The quality of primary data used in the implementation of the NN strongly influences the training phase, for this reason it plays a crucial role in the achievement of good results. To learn the features of the object the machine needs a significant level of data entropy. This was easily reached in the case of the Tour Saint Jacques, since in addition to historical images retrieved for both case studies, it was also possible to collect hundreds of contemporary images with different backgrounds, lighting conditions and points of view. The following methods were used for the collection: (1) web crawling; (2) ad hoc photographic survey at the new location of the Tower; (3) consultation of the historical archives in Paris.

The experiment was carried out in two different phases and with three different datasets, which are described in the following sections.

Dataset 1: Reference Case - Tour Saint Jacques

The first dataset was created to analyse two different Neural Networks on the best possible scenario of an existing heritage, such as the case study of the Tour Saint Jacques. The collected images of the Tour Saint Jacques (Figure 33) were first divided into four categories according to the following criteria:

1. Contemporary and historical images of the entire tower;
2. Views with the Parisian skyline, because they appear in the film footage;
3. Images showing monuments or architecture similar in shape (e.g. other towers) or style (e.g. Gothic architecture) to the Tour Saint Jacques. The latter images act as a "negative matching" and can reduce the incidence of false-positive ratio in classification problems of Deep Learning (Hu et al., 2014; Kalal et al., 2010);
4. Images that show only details or parts of the tower, because this is a typical situation when dealing with frames where the camera moves by filming only parts of an object.

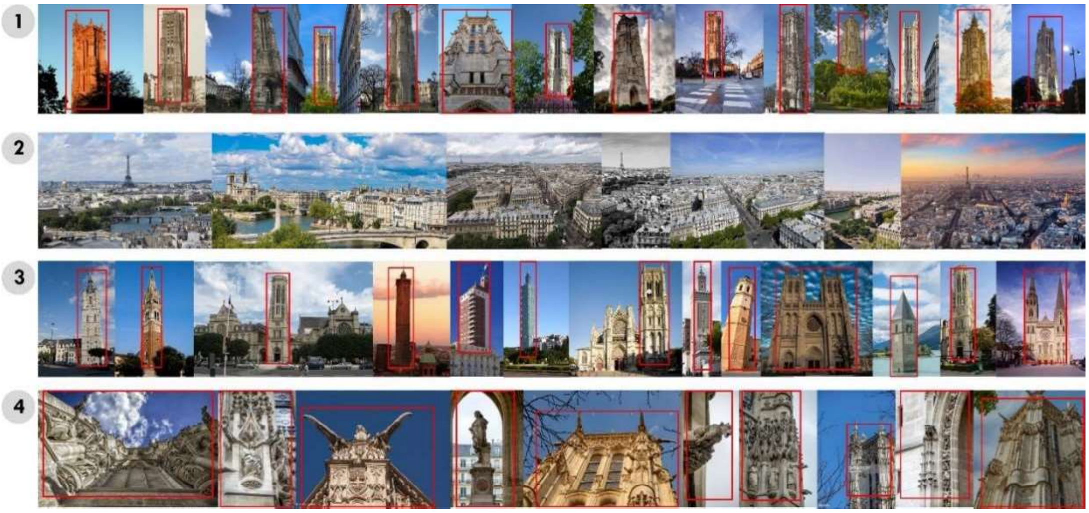


Figure 33. A selection of the pictures from the dataset 1: (1) Tour Saint Jacques; (2) landscape; (3) negative matching; (4) tower parts.

During the training phase, several combinations of number and type of images were extracted to improve the network performance, as shown in Table 1 and explained in the next section. In addition, during validation, 80 images from each group in the dataset, designated as valid1, valid2, valid3, and valid4 respectively, were used to evaluate the quality of the results from different perspectives (Table 1)

Table 1. Description of the training and validation dataset 1 on the reference case of the Tour Saint Jacques. For each training implementation (RUN A, RUN B and RUN C) different number and combination of images are used.

Description	**From web**	**From survey**	**From historical photographs**	**Number in training**			**Number in validation**
				RUN A	**RUN B**	**RUN C**	
Tour Saint Jacques	x	x	x	400	400	400	80
Landscape	x						80
Negative matching	x		x		200	200	80
Tour Saint Jacques Parts	x	x		80	80	80	80

Dataset 2: Video

The second dataset was created to test the performance of the algorithm in a realistic case. For both case studies, historical videos from the Paris archives were collected. Despite the critical nature of the exploitation of these materials (see Chapter 2), a large number of videos were collected, whose characteristics are described in Table 2 and Table 3 in which for each film footage the metadata available were reported.

Table 2. Description of the video dataset of the Tour Saint Jacques.

Dataset	**Duration**	**Year**	**Director**	**Type**	**Film**	**Colour**	**Archive**
La tour Saint Jacques	9min 47s	1967	J. Sanger	documentary		B&W	Ina.fr
Études sur Paris	76min	1928	A. Sauvage	documentary	16 mm	B&W	CNC and VOD
Paris, Roman d'une Ville	49min	1991	S. Neumann	documentary	16 mm	B&W	Forum des Images
Paris 2ème partie	4min 44s	1935	G. Auger	documentary	16 mm	B&W	Forum des Images

Passant par Paris	13min 39s	1955	P. Perrier	fiction	8 mm	B&W	Forum des Images
Vue Panoramique sur Paris	2min	1954	A. Lartigue	documentary	16 mm	B&W	Forum des Images
Un film sur Paris	45min	1926	C. Lambert, J. Levesque	documentary		B&W	Lobster
La nouvelle babylone	24s	1929	L. Trauberg, G. Kozintsev	historical		B&W	Lobster
Paris, 1946	13min	1946	J.C. Bernard	documentary		Colour	Lobster
La grande roue	4min 20s	1913		documentary		B&W	Lobster
Paris et ses monuments	7s	1912	Pathe	documentary		B&W	Lobster

Table 3. Description of the video dataset of Les Halles.

Dataset	**Duration**	**Year**	**Director**	**Type**	**Film**	**Colour**	**Archive**
Crainquebille	1min32s	1922	J. Feyder	drama	35 mm	B&W	Lobster
Les Halles 1960	35min15s	1960		amateur		Colour	Lobster
Paris Mémoire d'écran	21s			documentary		B&W	Gaumont Pathé Archives
Le ventre de Paris	5min55s			documentary		B&W	Ina.fr
Le ventre de paris	3min11s	2008	JP. Beaurenaut	documentary		Colour	Ina.fr
Les Halles centrales	12min 29s	1969	J. Sanger	documentary		B&W	Ina.fr
La Destruction des Halles de Paris	3min28s	1971	H. Corbin, J. Humbert	documentary	35 mm	B&W	Les Documents Cinemato-graphiques
Le dernier marché aux Halles de Paris	2min28s	1969	G. Chouchan	documentary		B&W	Ina.fr
Les Halles : histoire d'un marché incontournable à Paris	2min5s			documentary		B&W	Ina.fr

Les Halles de Paris en 1971	1min2s	1971	documentary	B&W	Ina.fr
Les Halles	2min37s		documentary	B&W	Ina.fr

Dataset 3: Real Case—Tour Saint Jacques and Les Halles

The evaluation of the implementation of NN on film footage in which a lost monument appears means that it is not possible to use a dataset with contemporary images of the building, since it was destroyed. For this reason, the third dataset (Figure 34 and 35) was created to test the algorithm of this real situation on the two case studies. For this purpose, the image categories were divided into three different groups:

(1) Historical photographs of the monument.

(2) Historical images, both photographs and images extracted from the video dataset in which the searched monument appears.

(3) Negative matching, for the Tower this coincides with the third group of the first dataset; for the Les Halles it is the images with the buildings in Paris that appear in the film.

In addition, for the Tower, some images were collected in a new validation group called valid5 and added to the previous dataset 1 to test the algorithm on this reference case. For Les Halles, the validation group on which the algorithm was tested is called valid1. The number of images used during training and validation and the combinations for the different runs are listed in Tables 4 and 5 and are explained in more detail in the next section.

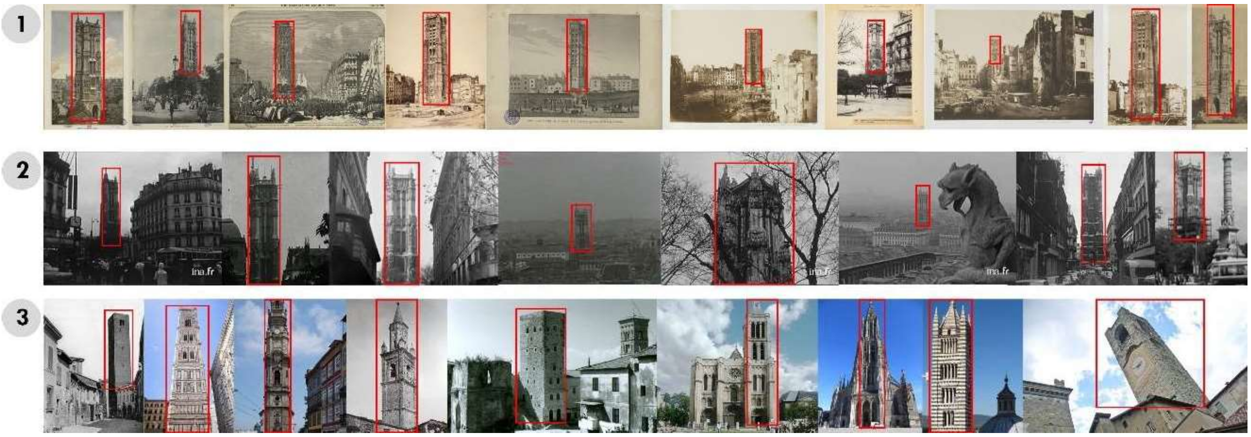


Figure 34. A selection of the pictures from the dataset 3 for the tower: (1) historical photographs; (2) historical images; (3) negative matching.

Table 4. Combination and number of images from the dataset 3 used in each run of the training and validation phase on the real case of the Tour Saint Jacques.

Dataset	Number in training RUN												Number in validation
	T1A	T1B	T1C	T1D	T1E	T1F	T2A	T2B	T2C	T2D	T2E	T2F	VALID5
1. Tour Saint Jacques – Historical Photographs	50	42	35	25	15	5	25	20	10	5	7	2	0
2. Tour Saint Jacques – Frame	0	0	0	0	0	0	25	22	25	20	8	3	29
3. Negative matching	50	42	35	25	15	5	50	42	35	25	15	5	0

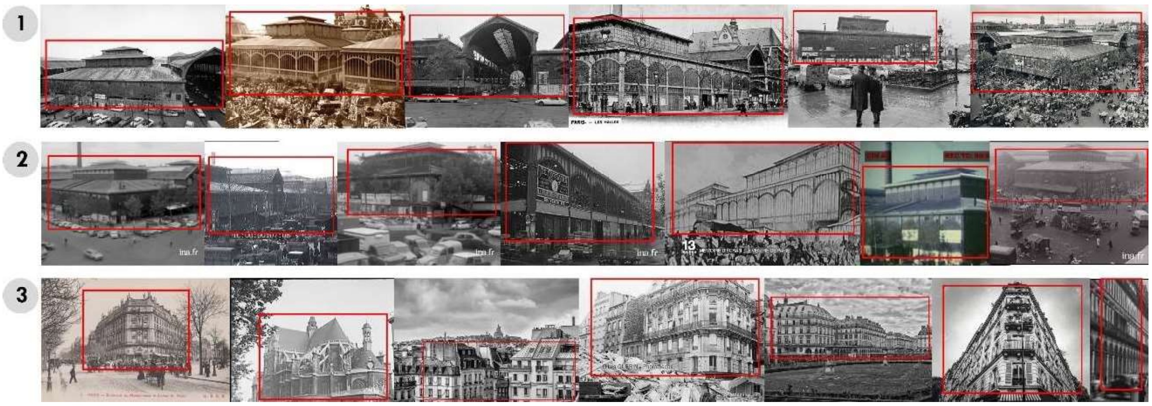


Figure 35. A selection of the pictures from the dataset 3 for Les Halles: (1) historical photographs; (2) historical images; (3) negative matching.

Table 5. Combination and number of images from the dataset 3 used in each run of the training and validation phase on the real case of Les Halles.

Dataset	Number in training RUN												Number in validation
	T1A	T1B	T1C	T1D	T1E	T1F	T2A	T2B	T2C	T2D	T2E	T2F	VALID1
1. Halles Historical Photographs	50	42	35	25	15	5	25	17	23	16	6	0	32
2. Halles Frame	0	0	0	0	0	0	25	25	12	9	9	5	46
3. Negative matching	50	42	35	25	15	5	50	42	35	25	15	5	0

4.2.2 Results and Discussion of the implementation of the Neural Networks on the case studies

This section deals with the results of the Neural Network referring to the metrics introduced in Chapter 3 and chosen for evaluating their performance. A sketch of the metrics chosen is reported in Figure 36 and examples of frames of True Positive, False Positive, True Negative and False Negative resulted from the detection of monuments by Neural Networks are reported in the following Figure 37-45.

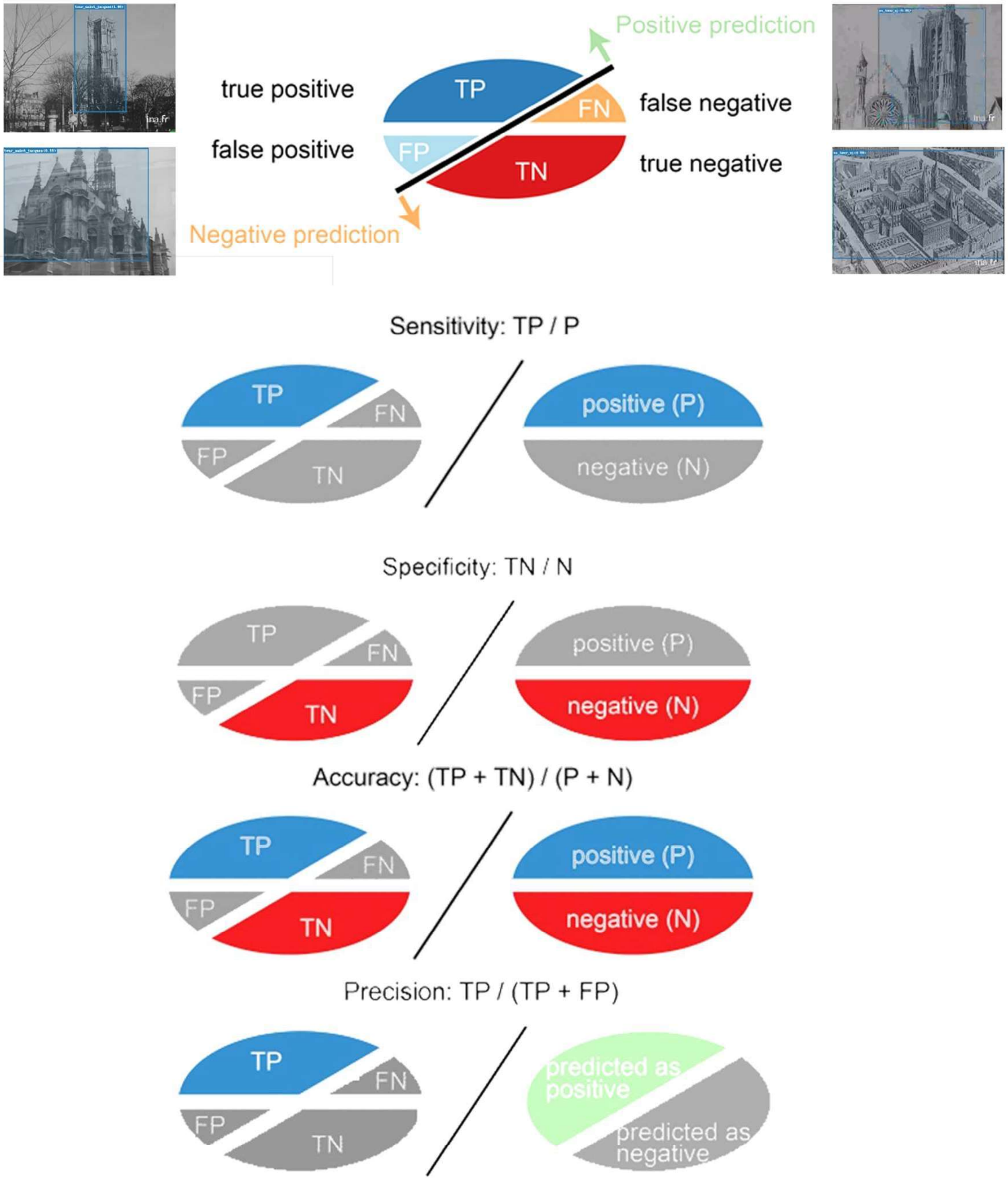


Figure 36. A sketch of the metrics chosen to evaluate Neural Networks.

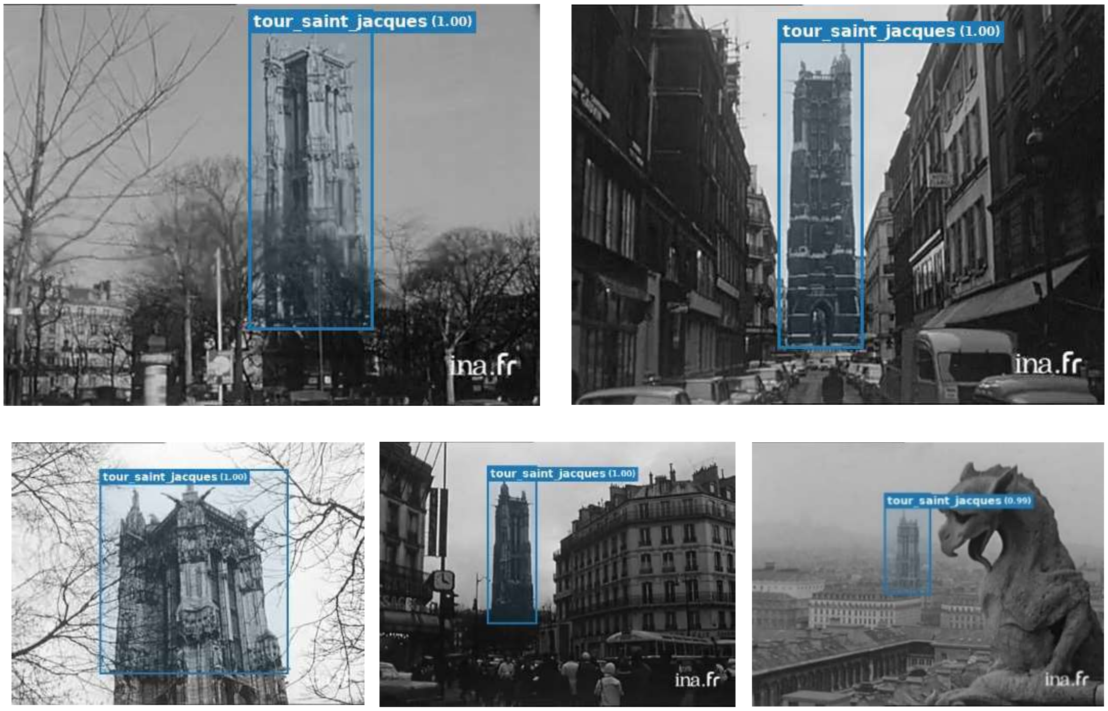

tour_saint_jacques (1.00)
ina.fr
tour_saint_jacques (1.00)
ina.fr
tour_saint_jacques (1.00)
tour_saint_jacques (1.00)
ina.fr
tour_saint_jacques (0.99)
ina.fr

Figure 37. Examples of frames True Positive of the Tour Saint Jacques: the monument was correctly detected by Neural Networks.

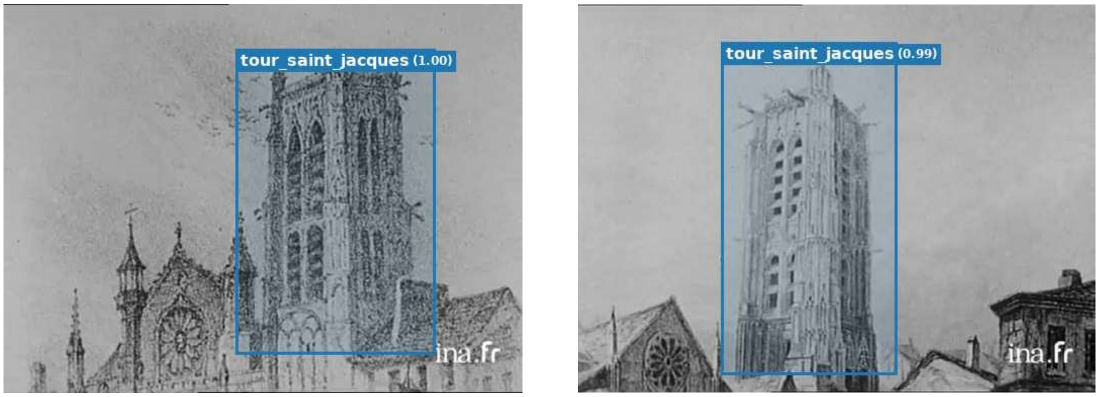


Figure 38. Examples of frames True Positive of the Tour Saint Jacques: the monument was correctly detected by Neural Networks, even if it is a drawing.

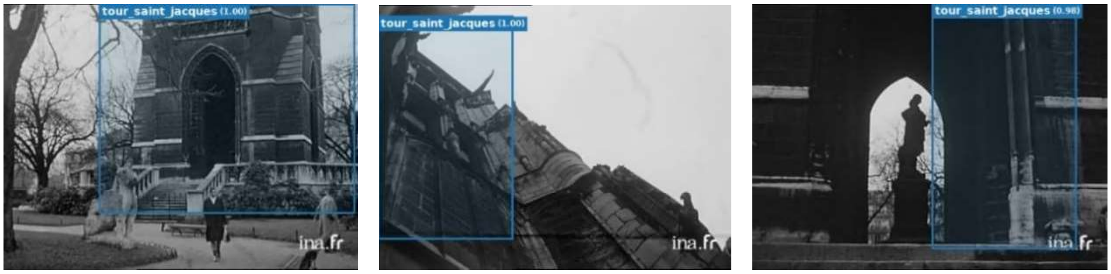

tour_saint_jacques (1.00)
ina.fr
tour_saint_jacques (1.00)
ina.fr
tour_saint_jacques (0.98)
ina.fr

Figure 39. Examples of frames True Positive of the Tour Saint Jacques: even the details and parts of the monument were correctly detected by Neural Networks.

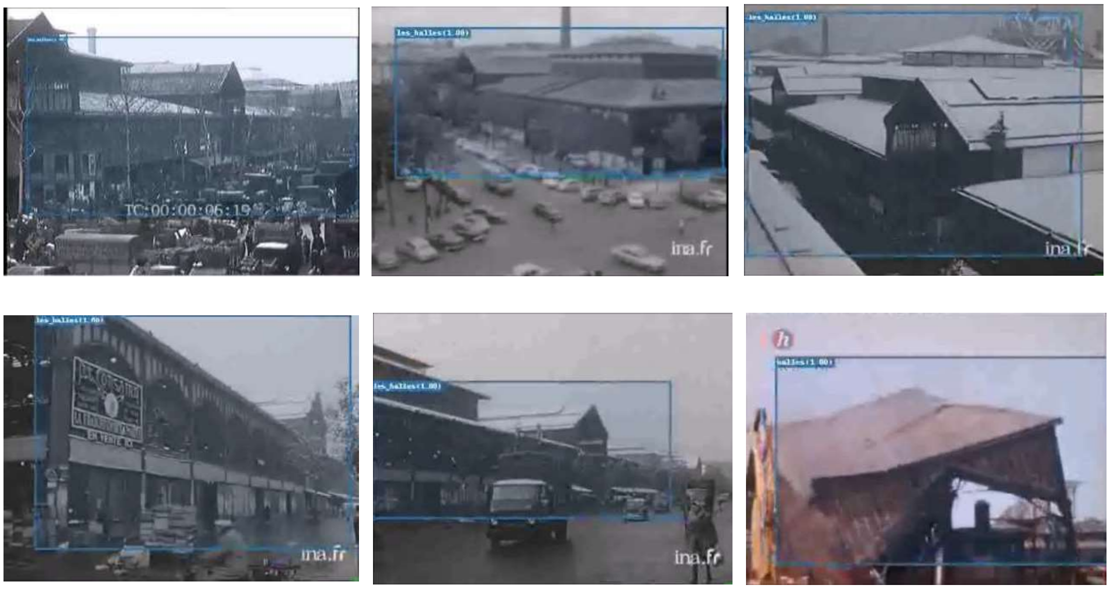


Figure 40. Examples of frames True Positive of Les Halles: the monument was correctly detected by Neural Networks.

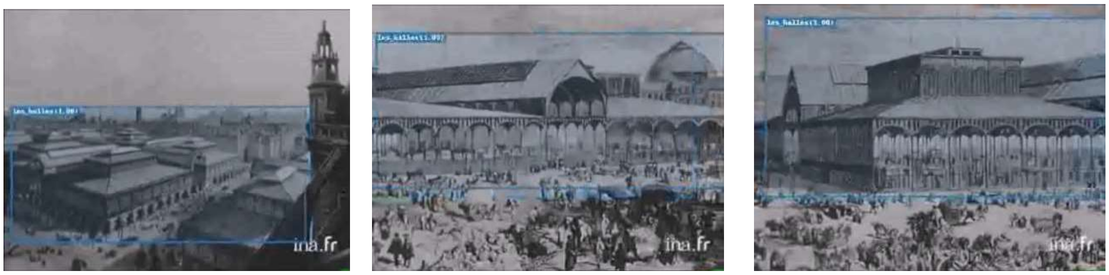


Figure 41. Examples of Les Halles: the monument was correctly detected by Neural Networks, even if it is a drawing.

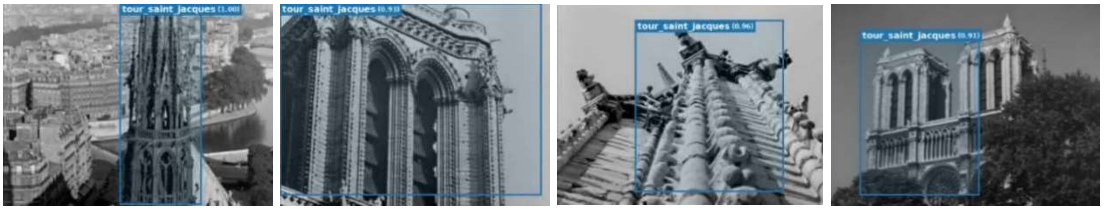

tour_saint_jacques (1.00)
tour_saint_jacques (0.93)
tour_saint_jacques (0.96)
tour_saint_jacques (0.91)

Figure 42. Examples of frames False Positive of the Tour Saint Jacques: the object was detected by Neural Networks as “Tour Saint Jacques” but it is not the tower.

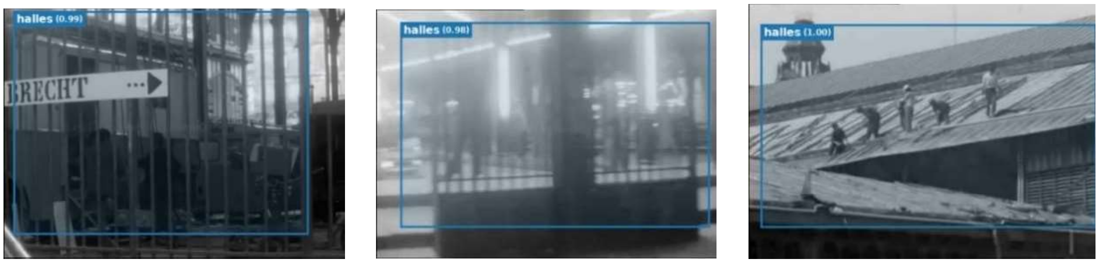
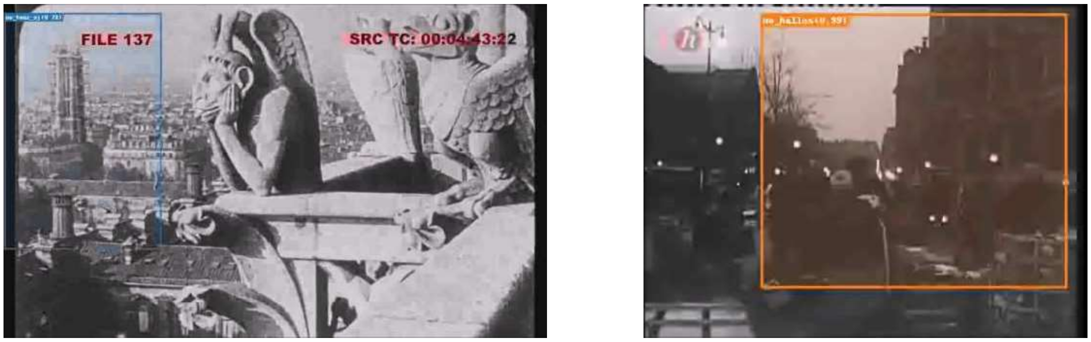
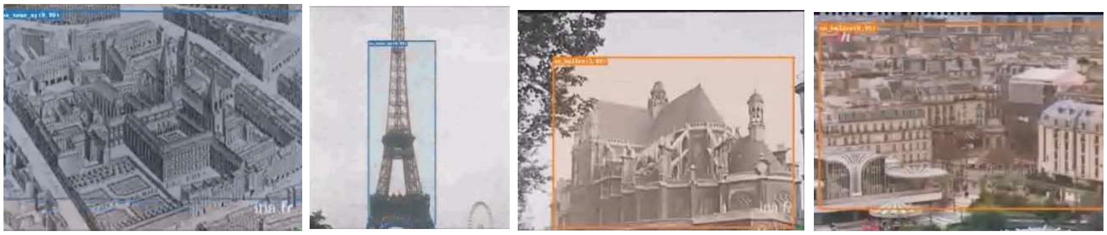


Figure 43. Examples of frames False Positive of Les Halles: the object was detected by Neural Networks as “Les Halles” but it is not the pavilion.



(a) (b)

False 44. Examples of frames False Negative of (a) the Tour Saint Jacques and (b) Les Halles: the objects were detected by Neural Networks as “No Tour Saint Jacques” and “No Halles” but they are the searched objects.



(a) (b)

False 45. Examples of frames True Negative of (a) the Tour Saint Jacques and (b) Les Halles: the objects were correctly detected by Neural Networks as “No Tour Saint Jacques” and “No Halles”.

The first part of this section describes the training phase and discusses the choice of the network model, the type of training dataset and the selection of the probability threshold. The second part discusses the influence of size and source of the data set on training. In the third subsection, the network is evaluated in a realistic scenario, focusing on the behaviour of the parameters that are most closely related to the activity of the end-user. In the fourth part a short discussion of the computational power required for the use of Neural Networks is presented. Finally, in the fifth part, the results of Neural Networks are discussed with respect to the last phase of the pipeline, the photogrammetric reconstruction.

Network Model Selection and Tuning

The first part of the training experiments is based on the case of the Tour Saint Jacques for the wide availability of past and modern images as well as some historical video sets. The experiments started with the assumption of the RCNN network, also called "accurate network" in the Luminoth reference. In the first training - called RUN A - only positive matches were used, represented by images of the Tour Saint Jacques with full or partial views (dataset 1 and 4). The results of the sensitivity analysis performed on the training and validation sets containing the tower (valid1 and valid4, respectively) are shown in Figure 46, taking into account two acceptance thresholds for the reference probabilities of 0.5 and 0.9, respectively.

What can be clearly seen in the graph is that the network converges quickly and achieves a very high training accuracy (i.e. sensitivity) as well as very high validation accuracy values for the valid1 set, which contains only images of the entire tower. The accuracy is more limited for valid4 (about 0.8) because partial views of the tower are not always recognized. In addition, the network is expected to become more selective by increasing the probability threshold and the sensitivity tends to decrease significantly, especially for valid4.

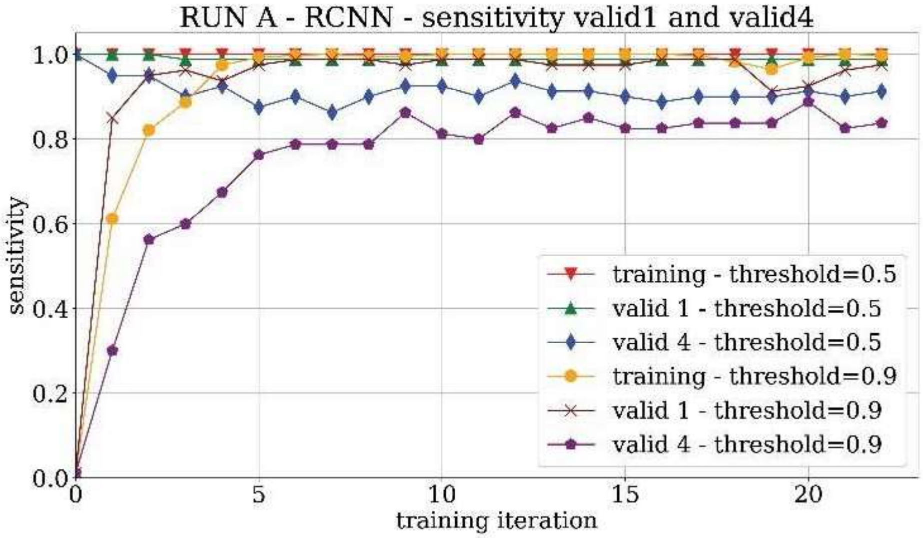


Figure 46. Sensitivity trend of RUN A with valid 1 and 4.

The validation behaviour of validation sets containing images without the tower is shown in Figure 37 for the valid2 and valid3 sets. For the valid2 set containing the images with the landscape around the position of the tower, the trend is quite good and reaches a specificity value of 90%. For the valid3 set, the presence of towers different from the Saint-Jacques leads to confusion in network learning and poor results close to 50% specificity. This means that the network is not able to distinguish the real tower from other similar towers with high accuracy. The higher the threshold value, the more selective the network becomes and the problem of false positives is thus at least partially mitigated.

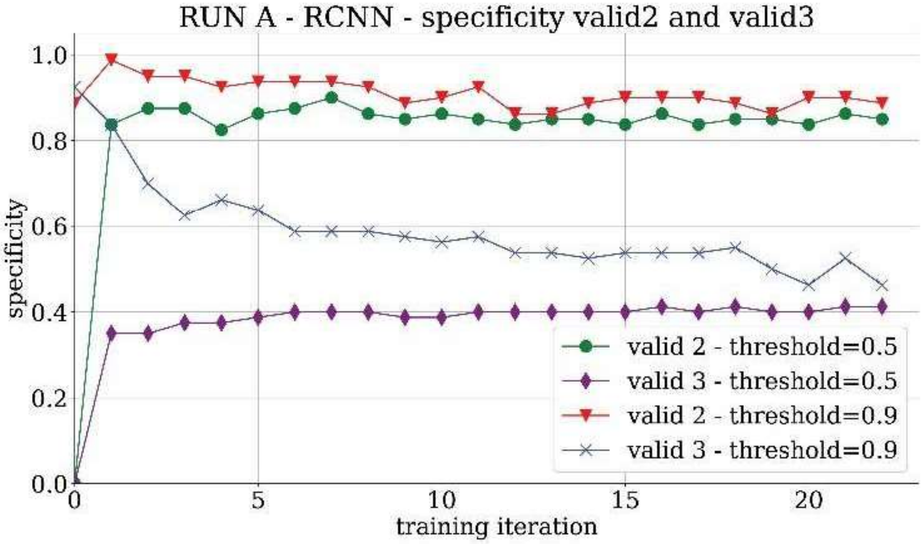


Figure 47. Specificity trend of RUN A with valid 2 and 3.

The difference between the specificity results of the valid2 and valid3 sets is not surprising. Since the training runs start from pre-trained networks, it is to be expected that the common categories - in the sense of Neural Networks - are already stored in the initial network weights. From the validation results shown in Figure 47 it can be seen that only a few false-positive results correspond to landscapes, while a considerable number of tower-like forms are misinterpreted by the network as Tour Saint Jacques. In order to deal with the most common type of false-positive result in this scenario, the set of images in valid3, which contained similar shapes to the searched Tour Saint Jacques, was used as a set of negative matching in subsequent runs.

The second training RUN B was still based on the Faster-RCNN network but it was performed including the negative matching set of images with the aim of improving the performance of the network minimizing the false-positive results.

Figure 48 shows the sensitivity analysis of the RUN B network. As expected, the network becomes more selective compared to the previous training scenario. The graphs reveal a slight degradation of the recognition of the true positives compared to RUN A where negative matching images are not used in training. However, in terms of specificity—as shown in Figure 49—the problem of false-positive results seems to be mostly solved. The significant improvement in specificity of RUN B compared to RUN A demonstrates that using valid3 as the “negative matching” set was an effective choice. Overall, the advantages of RUN B training outweigh the disadvantages. However, according to the use of the algorithm, it could be decided to always prefer sensitivity, so RUN A would be slightly better.

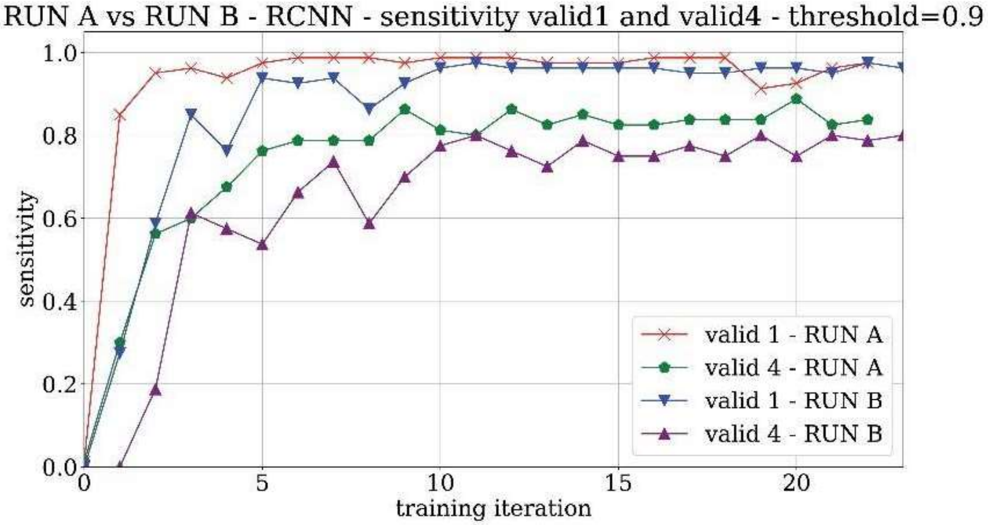


Figure 48. Sensitivity trend of RUN A and RUN B with valid 1 and 4, and threshold 0.9.

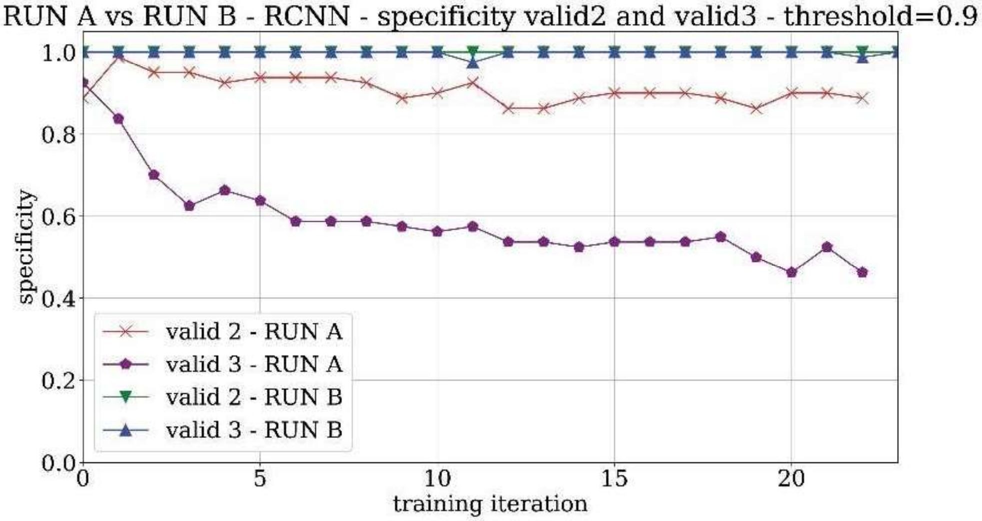


Figure 49. Specificity trend of RUN A and RUN B with valid 2 and 3, and threshold 0.9.

Figures 50 and 51 reveal the analysis of sensitivity and specificity trends for the two trainings RUN A and RUN B in order to assess the influence of the probability threshold on the results. The value of sensitivity is likely to decrease with the threshold whereas the specificity is expected to increase. From both figures, it results that 0.9 can be a good compromise on the threshold selection.

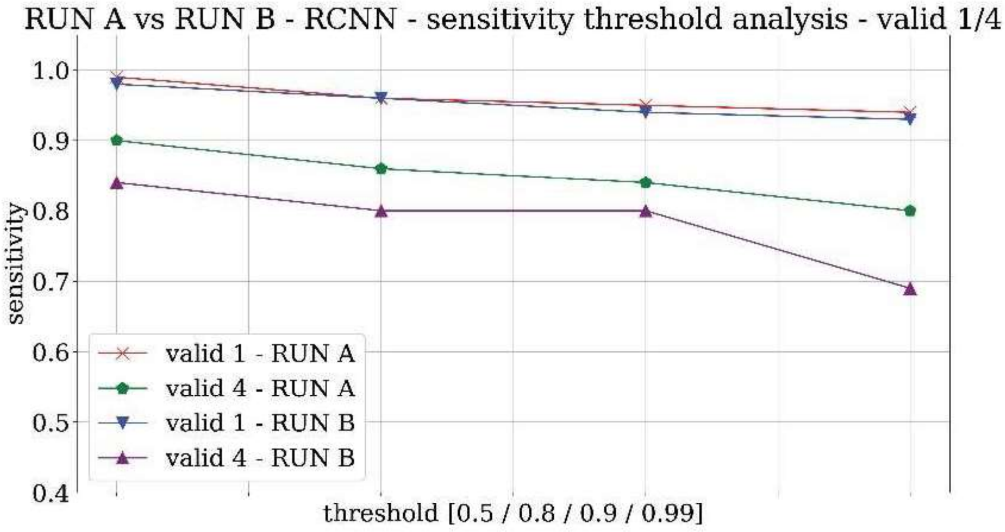


Figure 50. Sensitivity threshold analysis for RUN A and RUN B with valid 1 and 4.

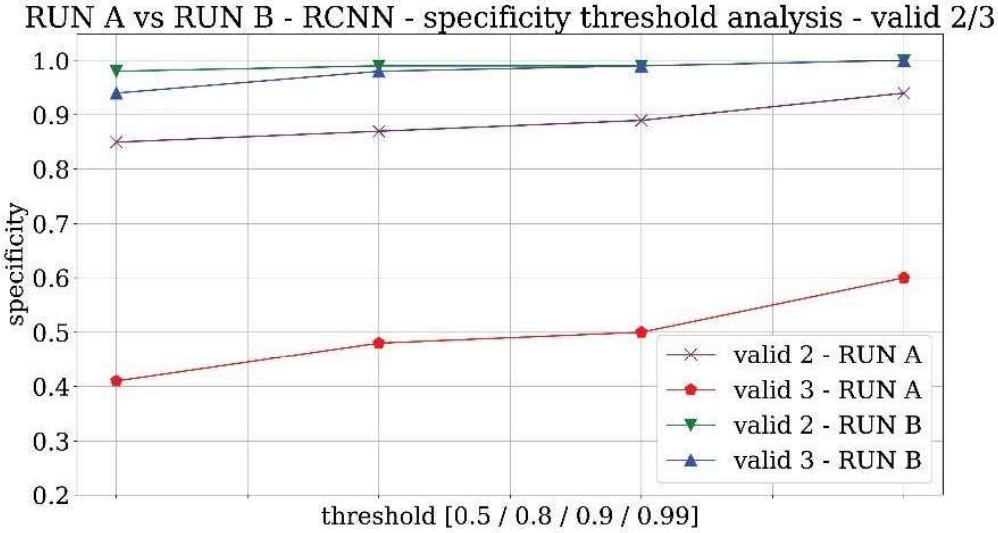


Figure 51. Specificity threshold analysis for RUN A and RUN B with valid 2 and 3.

The third training RUN C which was attempted is based on the SSD network, while keeping fixed the training dataset (including negative matches). The results of the comparison between the RCNN network and the SSD network are shown in Figures 52 and 53. What can be seen is that in positive cases the SSD network provides better values of sensitivity at least for the valid 4 set (Figure 52). However, from a visual inspection, it turns out that the images which are detected only by the SSD network are usually very poor-quality images or even drawings and therefore not suitable for the photogrammetric extraction. In terms of specificity, from Figure 53 it turns out that the values for both networks are high. Especially for the RUN B the values are very close to the ideal value of the unit. On the other hand, from Figure 53 it results that there is a specificity degradation for the SSD network. Even though, the specificity degradation seems small—around 2%—it is worth noting that in realistic scenarios the amount of negative images is huge and having 2% of false positives may be incredibly costly for the end-user activity.

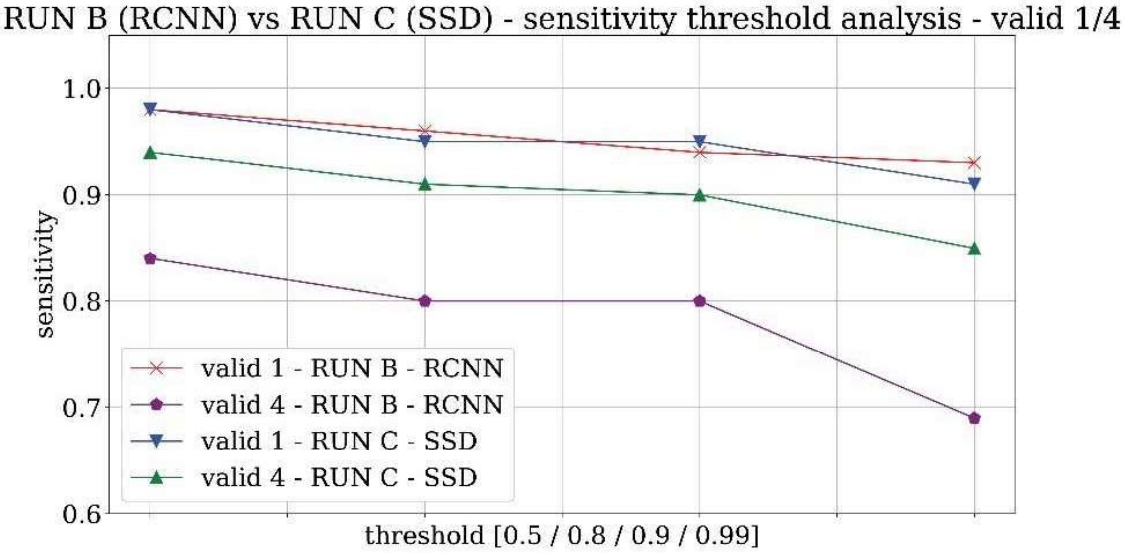


Figure 52. Sensitivity threshold analysis for RUN B and RUN C with valid 1 and 4.

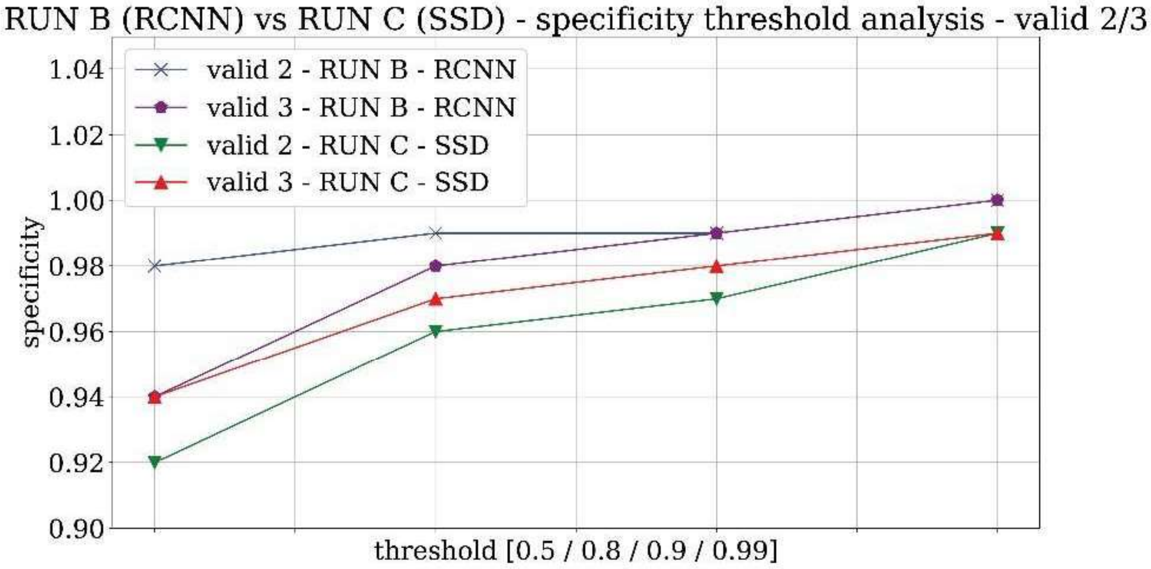


Figure 53. Specificity threshold analysis for RUN B and RUN C with valid 2 and 3.

In summary, the SSD network is lighter in terms of computation than the faster RCNN and can detect a larger number (usually of poor quality) of true positives, but at the same time detects too many false positives. The consequence is a larger amount of video to watch for the user. For these reasons, the threshold value of 0.9 together with the Faster-CNN were identified as the most reliable feature of the network to use in the next experimentations.

Assessment of the Training Dataset

The previous investigation identified Faster-CNN as a suitable model of the network to utilize. It also highlighted the advantage of inserting negative matching in the training datasets. Finally, the selection of 0.9 as the object detection threshold proved to be a good compromise. In the next part of this section, the application of these assumptions to a concrete case is examined. If the architectural heritage is lost, only historical archive material is available. For this reason, only historical images, both photographs and video images, were used in the following analyses. Given the difficulty of finding the first monument images necessary for the training, it is necessary to consider both the source of suitable images of the monument and the number of images required for a quality training phase. In order to study these aspects in the context of this study, several training phases were carried out. First, two different types of runs were selected. In the first case, the training was conducted using historical photos only, in the second case, images from historical videos were added to the dataset. To enhance generality, both the case studies of the tower and of the pavilions were tested: the terms T1 and H1 refer to the runs with the training datasets that contain only historical photographs, respectively, for the tour Saint Jacques and Les Halles. The terms T2 and H2 refer to the runs that also employ the frames extracted from videos. Furthermore, with the aim of calculating the minimum number of images required to achieve acceptable training results of the network, six different runs were processed with a decreasing number of elements in the dataset: these runs are labelled as A-B-C-D-E-F. For example, T1A stands for the run trained using 50 true images and 50 negative matchings, considering only historical photographs. On the other hand, T2A stands for the run trained using 50 true images and 50 negative matchings, considering 25 historical photographs and 25 video frames containing the monument. The other letters are related to the number of training images as follows: B=42, C=35, D=25, E=15, F=5. A comparison among each case study was performed by validating the network against datasets that contained only historical images, called valid5 (valid1) for the tower (Les Halles) test cases. The evaluation was achieved considering single frames of the videos.

The results are provided in Figures 54-55, where the sensitivity is plotted against the number of training images for each test case and each source of training images (only historical photo or combination of historical photo and video frames). Since the involved test datasets always contain the searched monument, the sensitivity equals the accuracy.

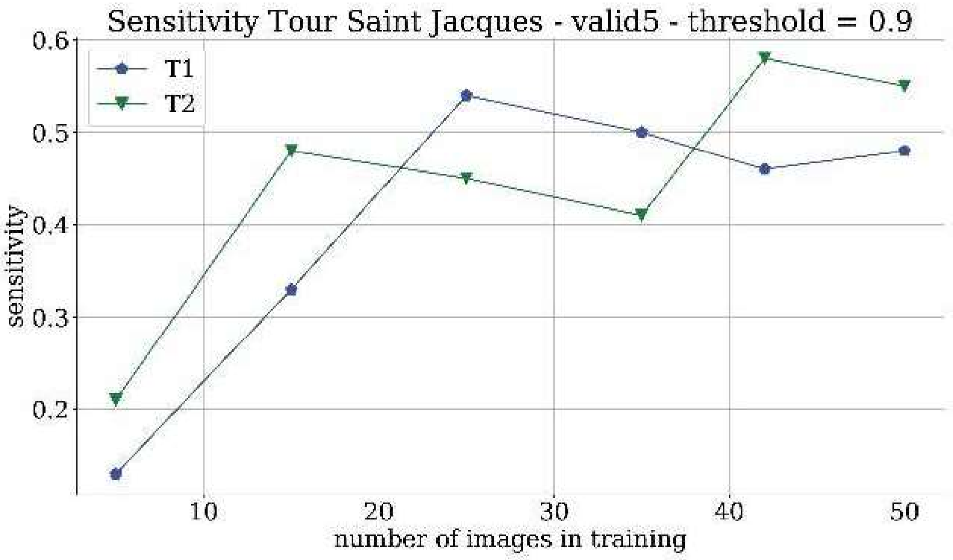


Figure 54. Real values of sensitivity analysis for T1 and T2 with valid5 and threshold equal to 0.9; (b) for H1 and H2 with valid1 and threshold equal to 0.9.

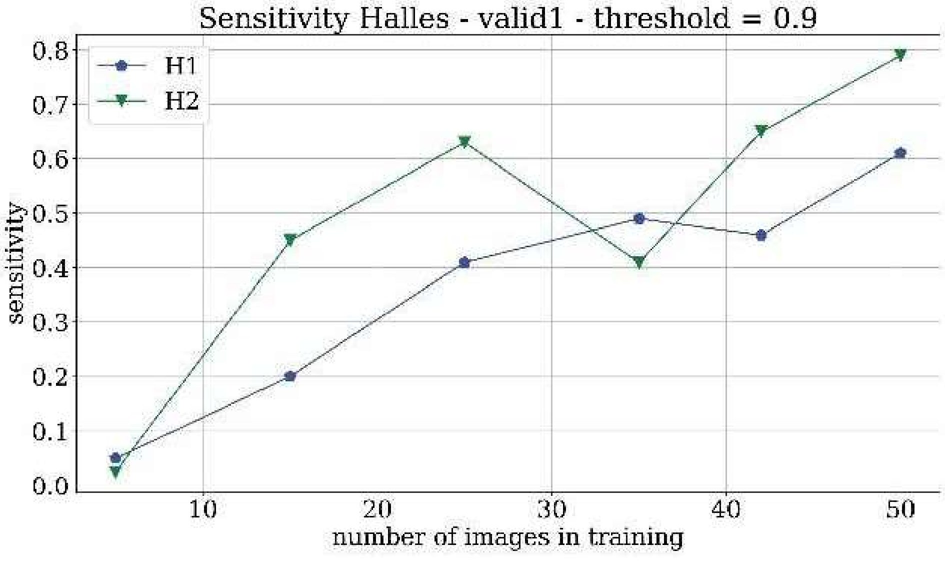


Figure 55. Real values of sensitivity analysis for H1 and H2 with valid1 and threshold equal to 0.9.

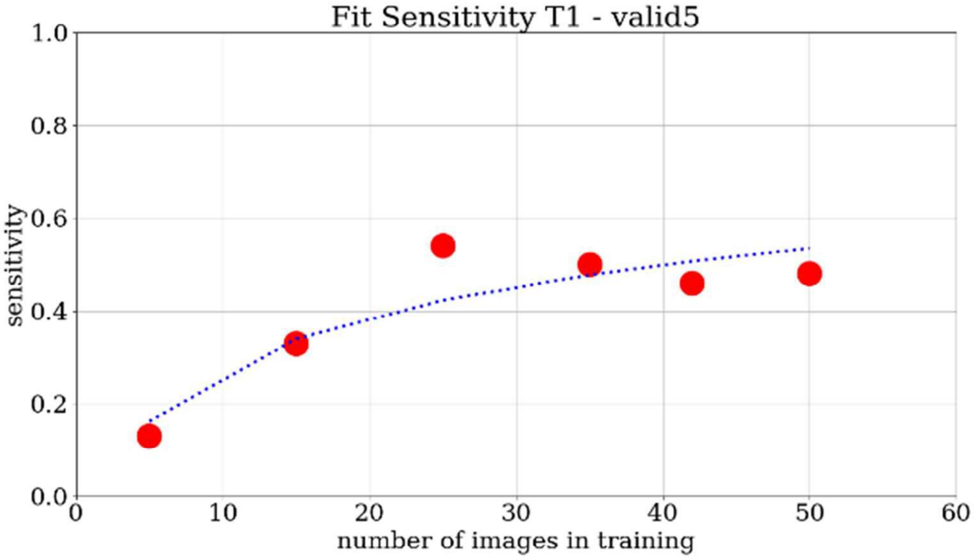


Figure 56. Real values (red dots) and fitting curves (dotted line) of sensitivity analysis for T1 with valid5 and threshold equal to 0.9.

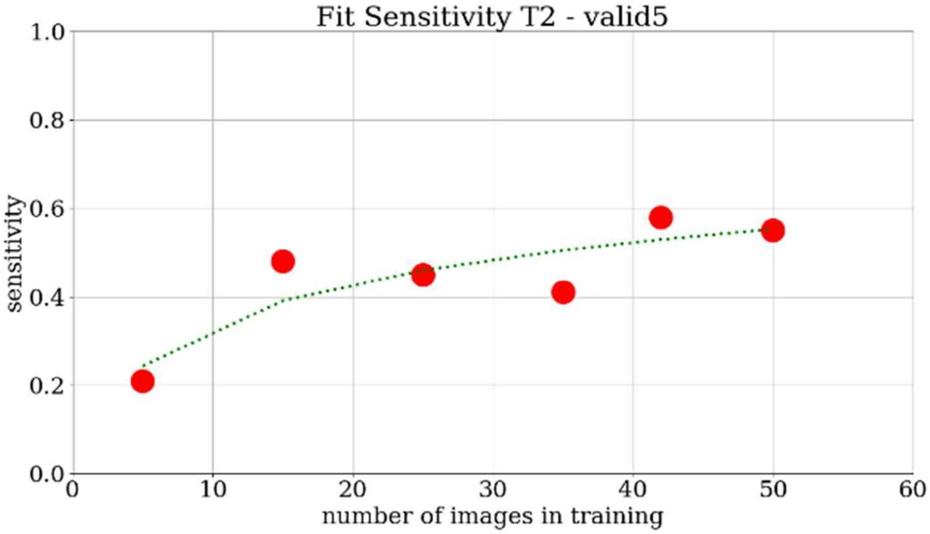


Figure 57. Real values (red dots) and fitting curves (dotted line) of sensitivity analysis for T2 with valid5 and threshold equal to 0.9.

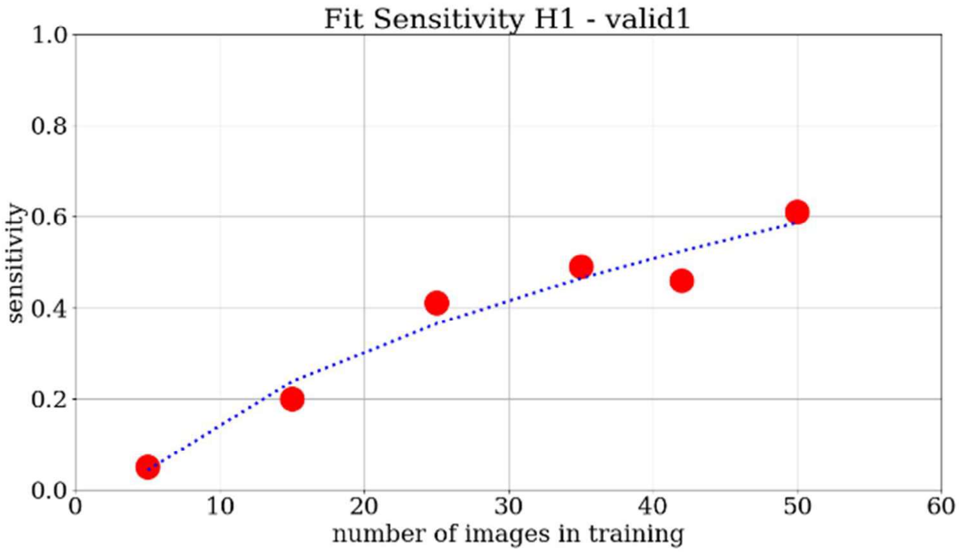


Figure 58. Real values (red dots) and fitting curves (dotted line) of sensitivity analysis for H1 with valid1 and threshold equal to 0.9.

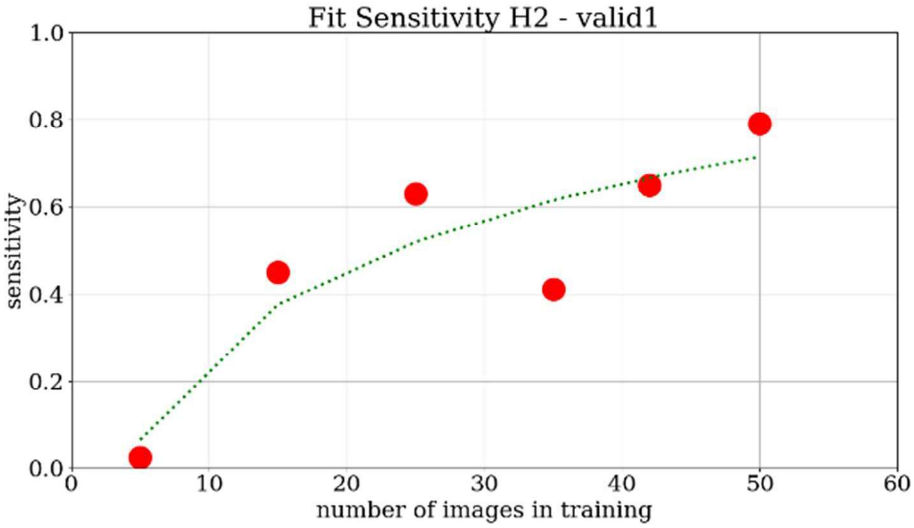


Figure 59. Real values (red dots) and fitting curves (dotted line) of sensitivity analysis for H2 with valid1 and threshold equal to 0.9.

The provided results show a monotonically increasing trend of the sensitivity when the number of images increases, for all of the four considered evaluations. Furthermore, a saturation trend for all of the cases can be ascertained. To achieve reasonably saturated results, around 25 images were required for the tower case, whereas 35 images were needed for the Les Halles case. What is interesting in these charts is that there is no great difference between the trend of the four curves, when compared to the curve fittings based on the simple power-law f(x) = a+b·xc. This is represented in

Figures 56-59. Even though this analysis is limited to two test cases, a first brief indication is that with a minimum of 30 images it is possible to train the network adequately to find the requested object. For low numbers of images, in particular the T2 and H2 cases present a measurable advantage in terms of sensitivity performances compared to the T1 and H1 counterparts. For a higher number of training images, the advantage is smaller and more difficult to detect. All in all, at least for these two test cases, the source and the type of the images do not significantly influence the performance of the learning process of the network. Instead, the number of training images is crucial to achieving good quality training.

Network Evaluation in a Realistic Scenario

Frame-Based Metrics

In the previous section, a detailed study of sensitivity performance with a variable number and source of training images was presented. The sensitivity essentially synthesizes the ability of the network to recognize the searched monument, but does not provide information about the time savings that can be achieved with neural networks compared to manual procedures. To discuss this last point, the first variable to be evaluated is specificity. In order to evaluate meaningful specificity values, a realistic test dataset is recommended. Indeed, in a realistic scenario where the amount of positive and negative is as balanced as expected in a real archive, it is possible to evaluate the best compromise between the metrics to be maximized. For this reason, the same analysis varying the number of training images is repeated considering the videos as the test dataset (as usual, the video frames used during the training are not used in the test dataset).

Starting with the use of the standard metrics applied to video frames, the resulting charts are provided in Figures 60 and 61 for both sensitivity and specificity parameters. With respect to sensitivity, the trends against the number of training images are similar to the previous sensitivity analyses, but the absolute values are slightly smaller, as expected considering that the average quality of video frames is lower than the historical photos used in the previous test datasets. All in all, the analysis shows that the monuments correctly identified as positive are less than half of the total positive ones, therefore some information is lost. In terms of specificity, the trends are not monotonous; the ability to detect true negatives fluctuates but does not show a definite trend in all four cases. The specificity values are always above 84%. This percentage refers to the time-saving advantage for the end-user, but a direct interpretation of the value in this sense is not obvious. Furthermore, it should be noted that the specificity value is limited by the fact that the test videos were manually selected to contain at least one occurrence of the target object, a circumstance that does not correspond to a realistic analysis of the archive.

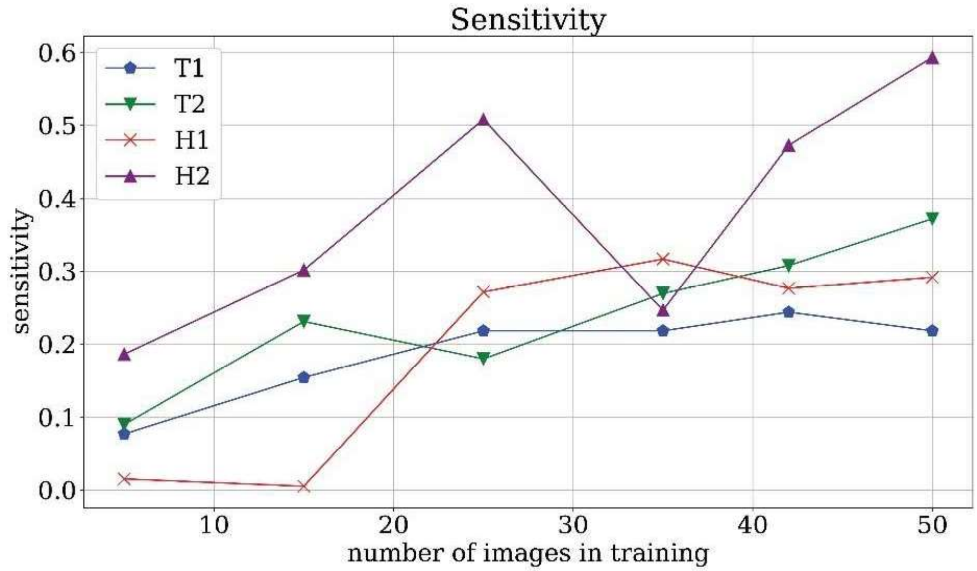


Figure 60. Sensitivity analysis evaluated on frames.

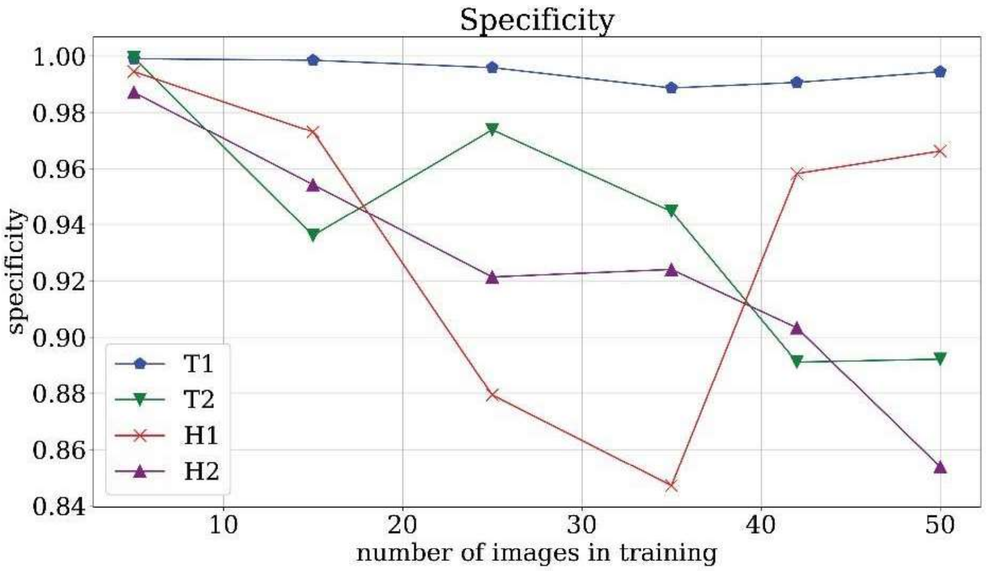


Figure 61. Sensitivity analysis evaluated on frames.

Time Interval-Based Metrics

As discussed in Section 3, the evaluation of metrics based on time-intervals may be more suitable to realistically analyse the quality of trained networks concerning the end-user activity.

In Figure 62 the time interval discovery-rate is plotted against the number of training images for T1, H1, T2, H2 cases. The percentage of correctly predicted intervals in which the monument appears found by the runs T2 and H2 of the network, in which both historical photographs and frames were used, reached a higher number than the T1 and H1 since the value of probability to detect the correct object is around 75% against 50%. As previously explained, the evaluation of the discovery rate is somehow related to the frame-by-frame sensitivity, even if calculated on intervals. Comparing the results of discovery-rate and standard sensitivity, it is evident that using a metric based on time intervals leads to an evaluation less strict than the counterpart based on the frame, but the time interval perspective is more significant from the point of view of the final user.

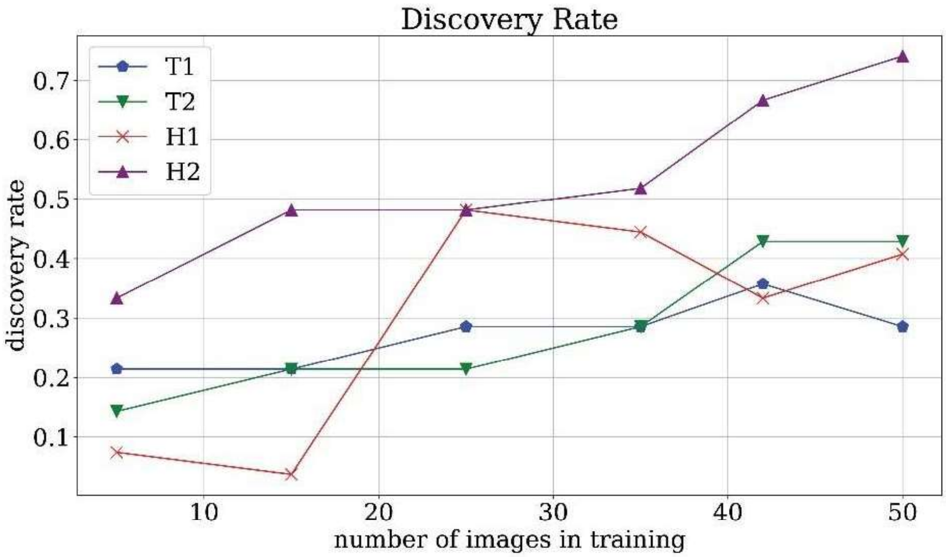


Figure 62. Discovery Rate analysis.

In Figure 63, the time save parameter is plotted against the number of training images. It turns out that for T2 and H2 runs with a low number of images the time save is around 1000. However, in this range the discovery rate is very poor. With the increase in the number of images the value decreases around 10/50 which is still a satisfactory value for the time saved. It is expected that the value increases, even more, when generic video archives are taken into consideration.

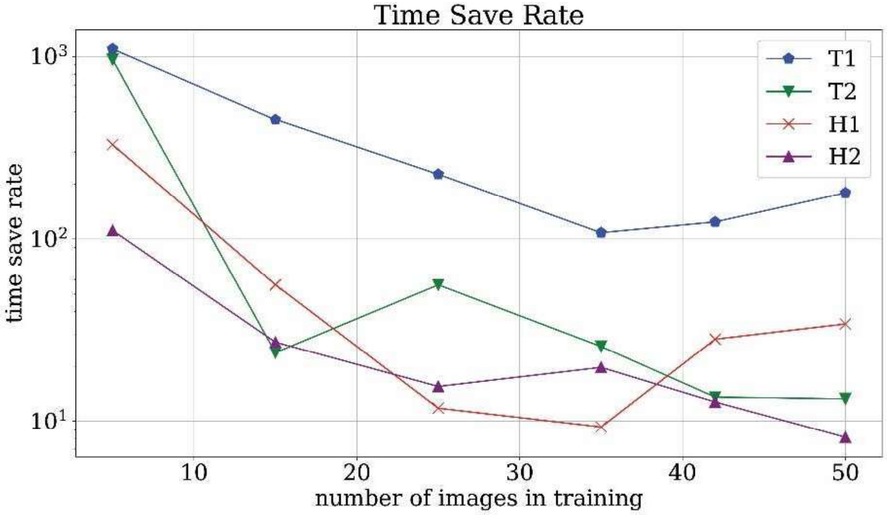


Figure 63. Time Save Rate analysis.

In order to get rid of the dependency on the type of considered videos, it is possible to compare the time save rate with the ideal time save rate, thus defining the time-saving efficiency. Time save efficiency results are plotted in Figure 64. For a low number of training images, the time save efficiency is higher than unity and this is due to poorly trained network, which is not capable of detecting both true and false positives. For mid-range and high-range numbers of training images, the efficiency is order 1 which means that the operator time save is close to the ideal time save. Obviously, this is possible because not all of the objects are correctly found. However, the found images are usually the best quality ones and therefore are more usable for the next steps of our pipeline. In this scenario, the time save rate efficiency close to unity can be considered an optimal result.

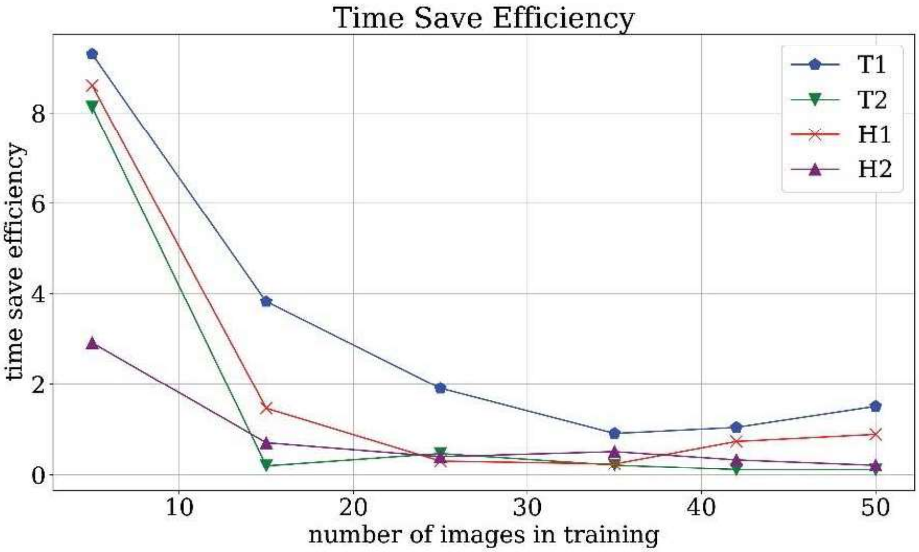


Figure 64. Time Save Efficiency analysis.

Hardware Analysis: High-Performance Computing vs. General Purpose

The High-Performance Computer of the Italian research centre CINECA (Rome, Italy) was used for the training and validation of Neural Networks) thanks to the award of ISCRA - Italian Super Computing Resource Allocation project.

The use of GPUs (Graphics Processing Units) has become a leading technology in the context of Deep Learning thanks to the high computing power available and the relatively low power consumption. Modern neural network frameworks support GPU computing. GPUs are available in ordinary home computers or small workstations, but GPUs are now also used as accelerators in high-performance computing (HPC) clusters. In Table 6 we show the time needed to process an image during the training phase.

To follow the evolution trend of the GPUs, results based on low/mid-range GPUs were reported up to results from top GPUs used in HPC centres. The order follows the release date of the devices. The type of GPU is also described as distinguishing HPC GPUs from consumer GPUs. It turns out that the improvement over the years is important, with a speed-up around 3 years after 5 years. Another very important point is the advantage of using HPC-oriented GPUs compared to normal laptop GPUs. The difference in timing is very marked. For complete training, the elapsed time may pass from several tens of days to less than 24h. In the massive inference phase, the use of HPC platforms can become a fundamental requirement.

Table 6. Hardware comparison.

NVIDIA 630M	**NVIDIA K40**	**NVIDIA P100**	**NVIDIA V100**	**NVIDIA 1650**
2012	2014	2016	2018	2019
Low-range Laptop	HPC	HPC	HPC	Mid-range Laptop
30s/image	1 s/image	0.5 s/image	0.3 s/image	9 s/image

4.2.3 Camera motions identification

In order to automatically select the frame sets to be used for photogrammetry, the algorithm for identifying camera movements described in Chapter 3, and reported again in Figure 65, was implemented. The procedure was tested on three different film footage, namely: “Tour Saint Jacques” from Ina.fr archive, “Études sur Paris” from the CNC-VOD archive and “La Destruction des Halles de Paris” from Les Documents Cinematographiques archive (see Table 2 and 3).

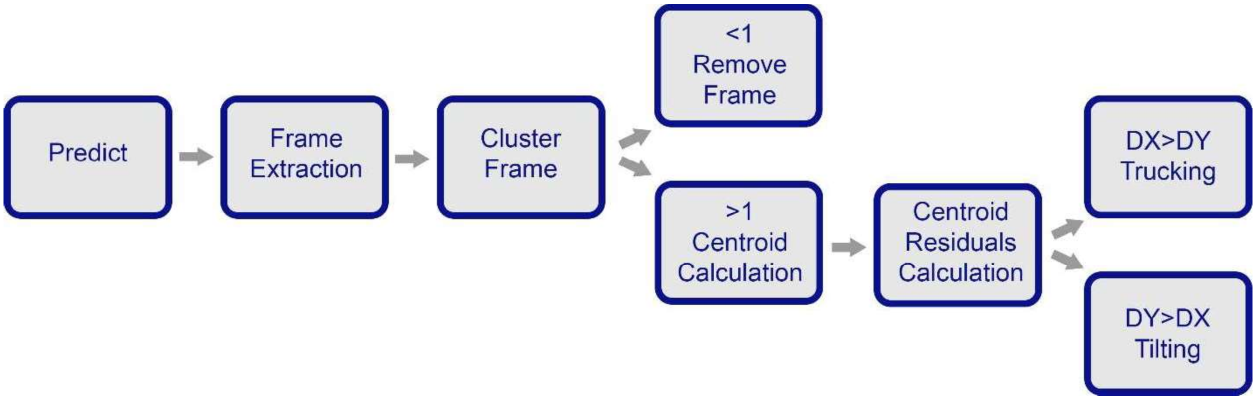
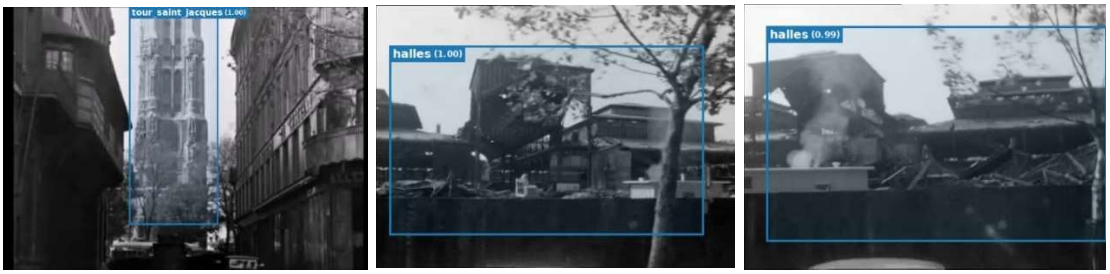


Figure 65.The workflow of the algorithm of camera motion identification.

The results of the predict step are the selections of the frames where the searched monument appears correctly detected by the Neural Network (Figure 66). The algorithm also identifies the name of the video and the timely appearance of the object within the video, together with the coordinates of the position of the bounding boxes in the frame and the probability score of the presence of the monument in the video. The frame extraction step provided the list of all images selected (Figure 67).



(a) (b)

Figure 66. An example of the frames resulted from the predict step in which the Neural Network correctly detected the (a) Tour Saint Jacques, and (b) Les Halles.

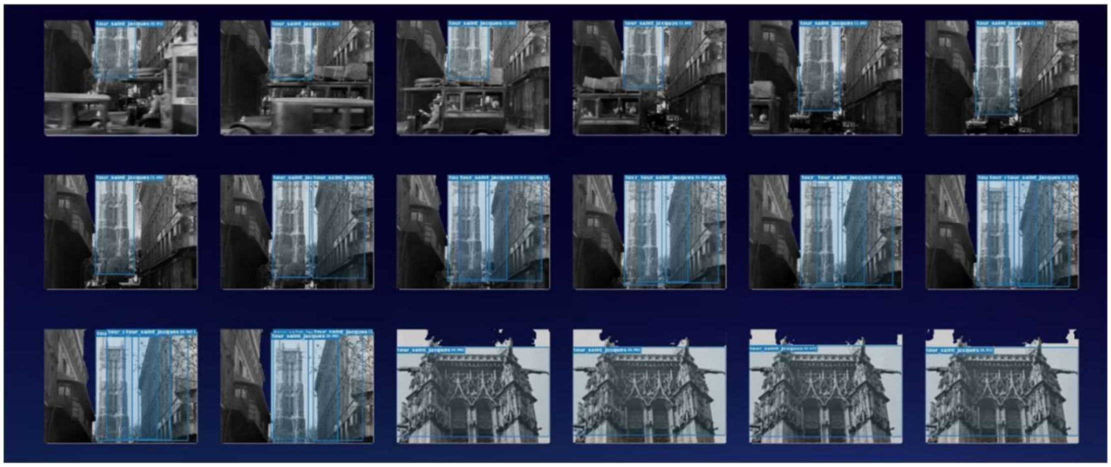


Figure 67. An example of the frames resulted from the frame extraction step in which the algorithm lists all the images selected by the predict step.

As explained in Chapter 3, the algorithm requires a correct selection of certain parameters. First of all, for the step of frame clustering of the algorithm, the second splitting criterion requires to adopt a structural similarity threshold Ts. Visual examination of the extracted frames revealed that setting Ts=0.1 is an effective choice for detecting a change of the scene in the great majority of the analysed cases.

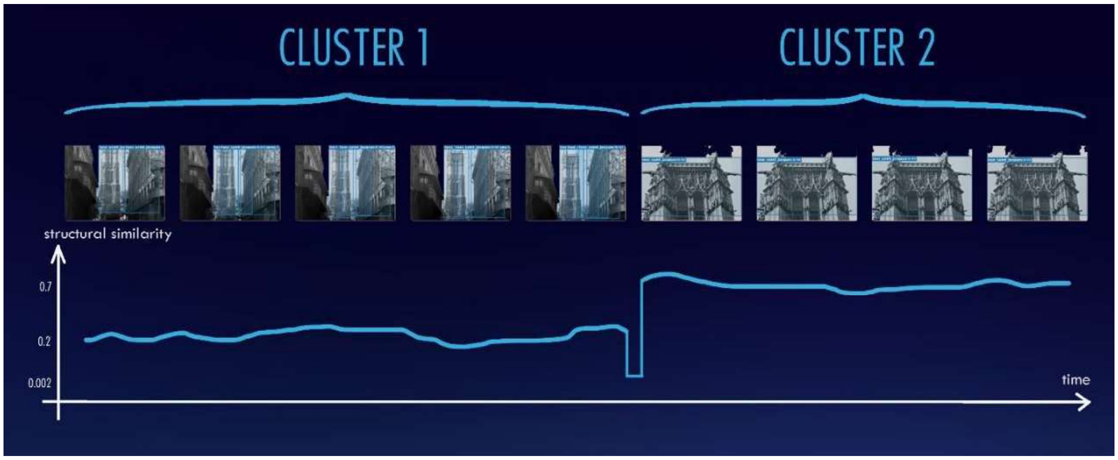


Figure 68. Example of the result of the frame cluster step according to the structural similarity threshold Ts. When the value of Ts is 0.1 there is a change of the scene.

During the camera motion estimation step, with respect to the Residual thresholds T1 and T2 (Figure 69), a selection is not straightforward and tuning based on tentative results seems to be a viable strategy. The results of the implementation of the algorithm on the frame selected are shown in Figure 70-73 following the categorization: "steady or oscillating camera", "camera trucking", "camera tilting" and "cannot determine camera motion".

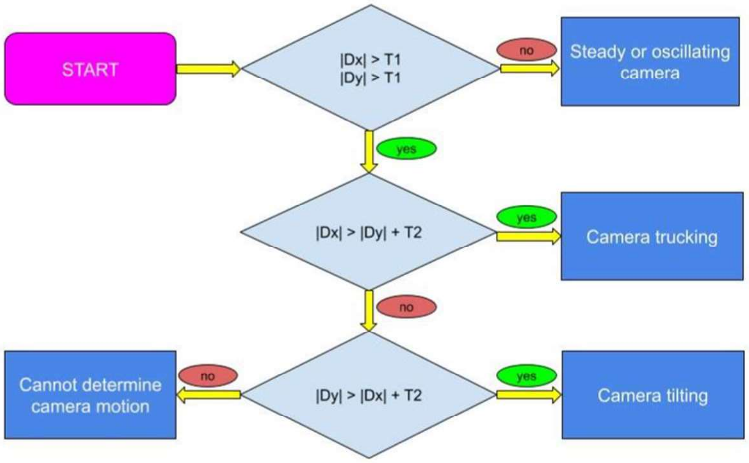


Figure 69. Workflow of the camera motion estimation step of the algorithm.

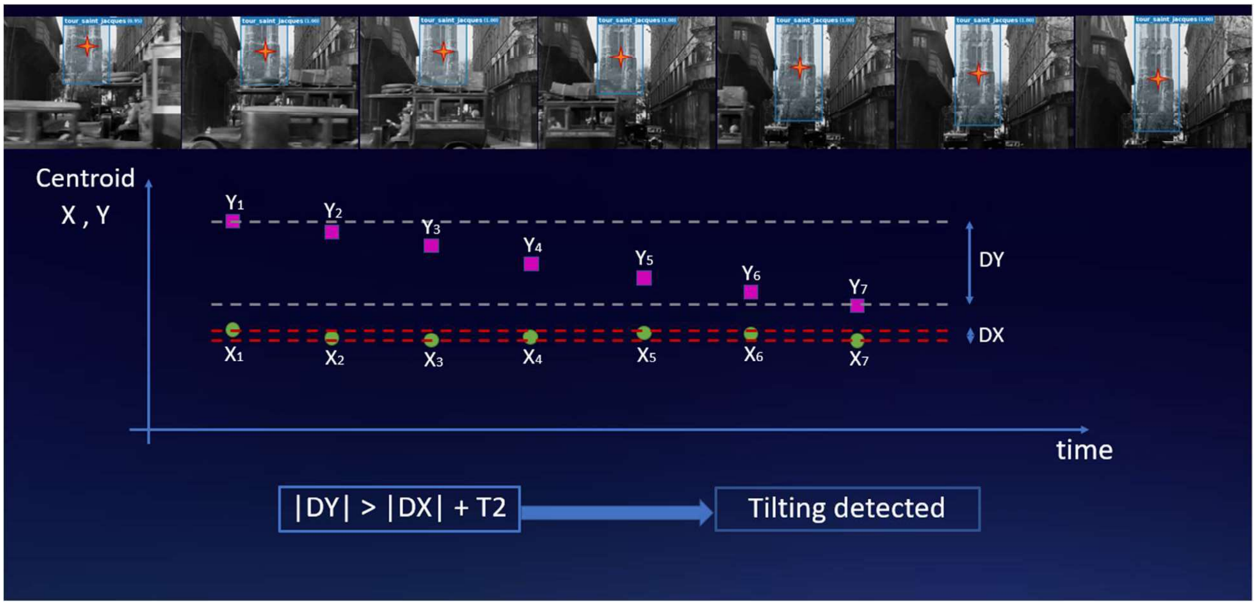


Figure 70. Results of the camera motion estimation step that recognised the " camera tilting" movement.

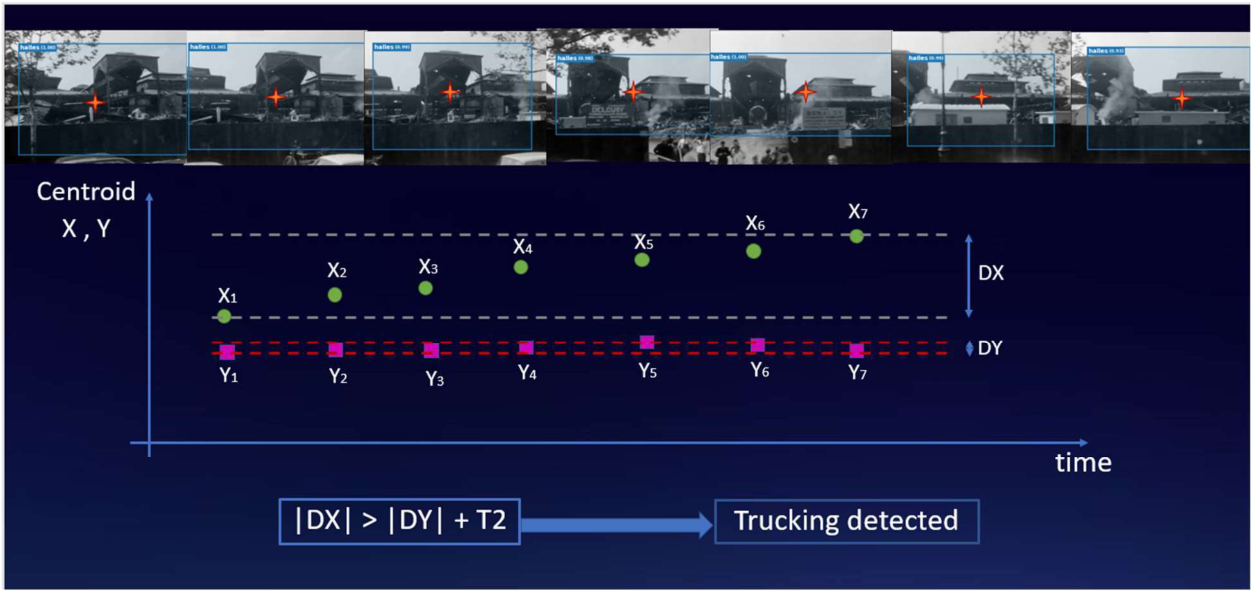


Figure 71. Results of the camera motion estimation step that recognised the "camera trucking" movement.

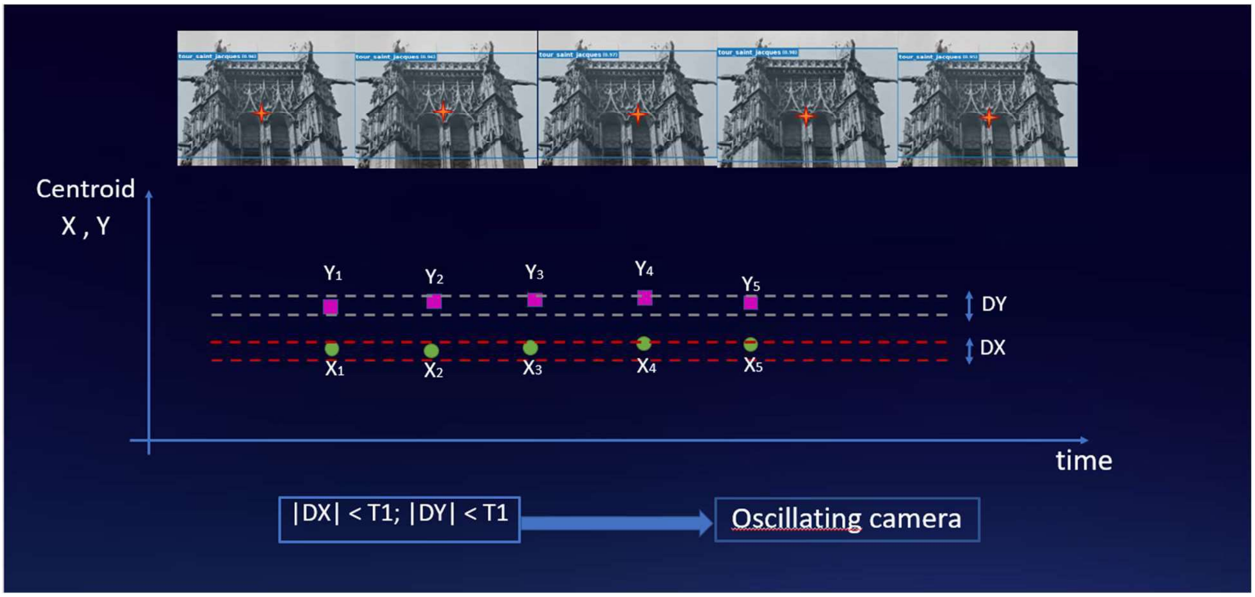


Figure 73. Results of the camera motion estimation step that recognised the “steady or oscillating camera” movement.

Table 7 shows the precision results for three different choices of (T1,T2), namely (0,0), (10,10), (20,20). To assess the accuracy, for each cluster of images, for each frame cluster the identified camera motion is compared against the real camera motion considering the four output categories, i.e. “steady or oscillating camera”, “camera trucking”, “camera tilting” and “cannot determine camera motion”. The accuracy is evaluated as the ratio between the number of correctly identified frame cluster motions and the total number of analysed frame clusters.

Table 7. Accuracy values expressed in percentage according to different clusters and thresholds.

Film	N° cluster	Thresholds 0.0 0.0		Thresholds 10.0 10.0		Thresholds 20.0 20.0	
		TOT [%]	NNTP [%]	TOT [%]	NNTP [%]	TOT [%]	NNTP [%]
Tour Saint Jacques	57	13	12	72	60	77	68
Études sur Paris	112	26	100	72	100	73	100
La Destruction des Halles de Paris	11	50	100	92	80	66	20

To allow a more detailed investigation of the results, the accuracy values are reported for each video separately. Moreover, for each video, two accuracy values are provided. The first one (TOT) is based on all the detected frame clusters while the second one (NNTP) only includes frame clusters which really represent the searched object. Indeed, since the camera motion algorithm is applied to the results of Neural Network (NN) object detection, some extracted frames are False Positive NN results, i.e. they do not correspond to the searched object. Since NN False Positives may correspond to other objects, or also to completely wrong image detections, it is expected that the camera motion algorithm will work more smoothly when filtering out these bad cases. In any case, summarizing, the TOT accuracy summarizes both NN and camera motion estimation accuracies, while NNTP accuracy more strictly refers to the camera motion algorithm accuracy.

The results in Table 7 show that when considering T1=T2=0, the TOT accuracy is poor (below 30%), whereas the NNTP accuracy is optimal for two of the videos but very poor for the third one. By increasing the thresholds to T1=T2=10, both the average accuracies are greater than or equal to 80%. Setting T1=T2=20, there is an accuracy degradation, especially for one of the videos. Concluding, the intermediate setup T1=T2=10 seems to be the best choice. All in all, it is clear that the number of analysed cases is not enough to demonstrate the effectiveness of the algorithm and of the entire workflow in a general context. However, the discussed preliminary results are encouraging. Not only the overall accuracy values are satisfactory, but the most suitable videos for photogrammetry are correctly identified and this means that, at least for the considered videos, the described workflow allows the user to reach the final goal which is identifying videos for photogrammetric reconstruction.

4.2.4 Photogrammetric processing and evaluation of metric quality assessment

Among the footage correctly detected by the Neural Network two different films to be processed were chosen. The first one is “Études sur Paris”, dated 1928, from the CNC-VOD archive (Figure 74) in which sequences of the Tour Saint Jacques taken with tilting camera motion appear. The second video, is “La Destruction des Halles de Paris”, dated 1971, found in Les Documents Cinematographiques archive in which the Pavilions appear shot with the trucking camera motion (Figure 75). The two films present the characteristics shown in Table 8.

Table 8. Technical features of the films used during the photogrammetric processing.

Film	**Gauge**	**Focal Length**	**Digital Format Resolution**	**N° Frame**	**Camera Motion**
Études sur Paris	16 mm	25 mm	480x360 pixels	16	Tilting
La Destruction des Halles de Paris	35 mm	35 mm	492x360 pixels	49	Trucking

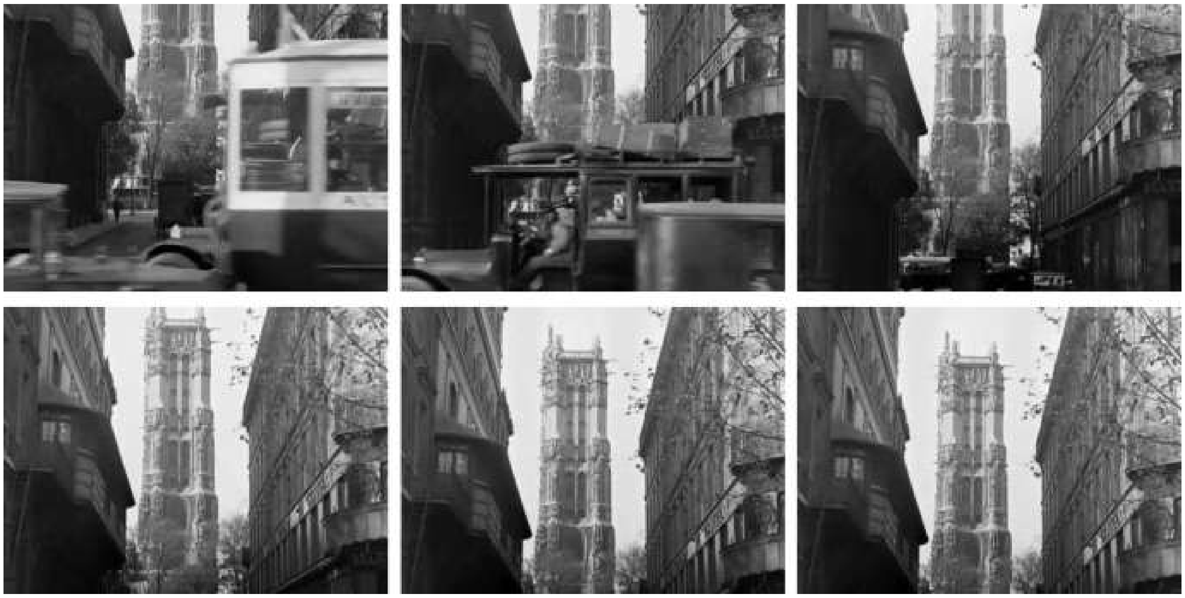


Figure 74. A selection of frames from the film footage "Études sur Paris" in which the Tour Saint Jacques appears shot with the tilting camera motion.

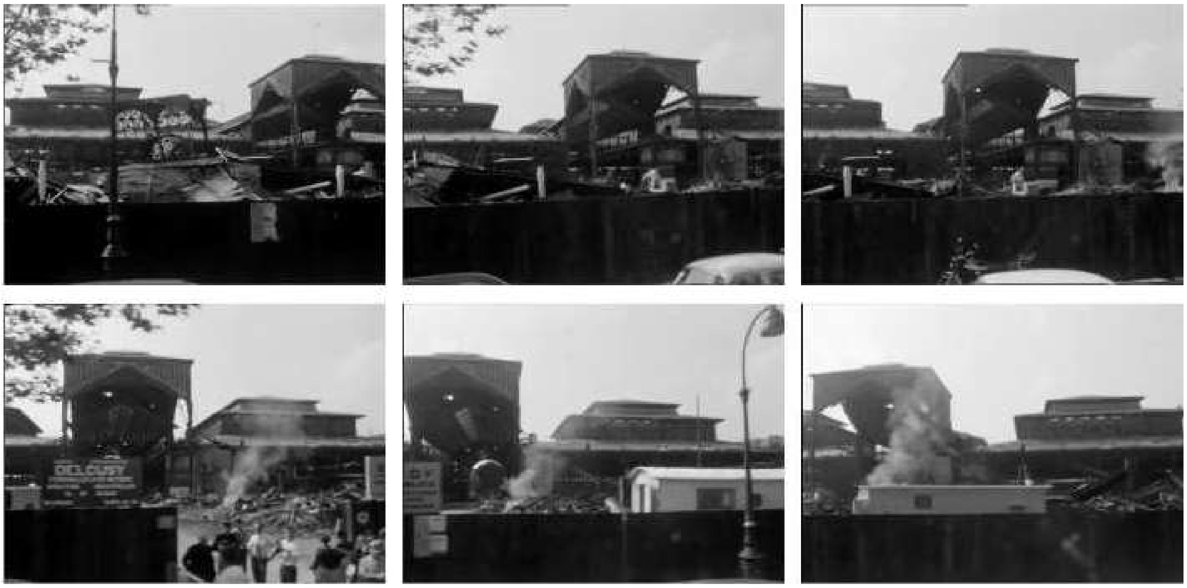


Figure 75. A selection of frames from the film footage "La Destruction des Halles de Paris" in which the Pavilion appears shot with the trucking camera motion.

Photogrammetric processing

According to the workflow proposed in this dissertation (Figure 76), the photogrammetric process implemented the SfM pipeline shown in Figure 77, referring to the COLMAP workflow. The results are reported in Figure 78 and 79.

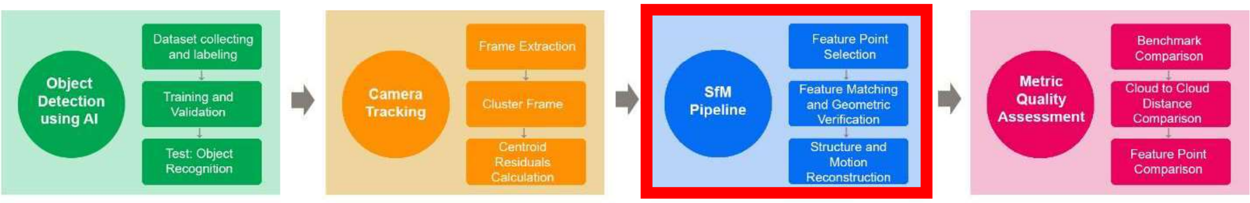

Object Detection using AI
Dataset collecting and labeling
Training and Validation
Test: Object Recognition
Camera Tracking
Frame Extraction
Cluster Frame
Centroid Residuals Calculation
SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction
Metric Quality Assessment
Benchmark Comparison
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 76. The general workflow proposed in this dissertation in which the third step concerning the photogrammetric processing is highlighted in red and reported below.

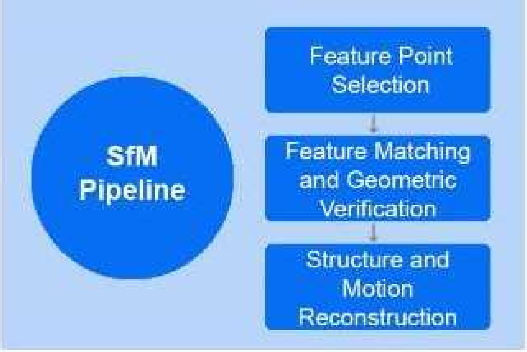

SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction

Figure 77. A focus on the third step of the general workflow proposed in this dissertation in which the three steps of the photogrammetric pipeline are reported.

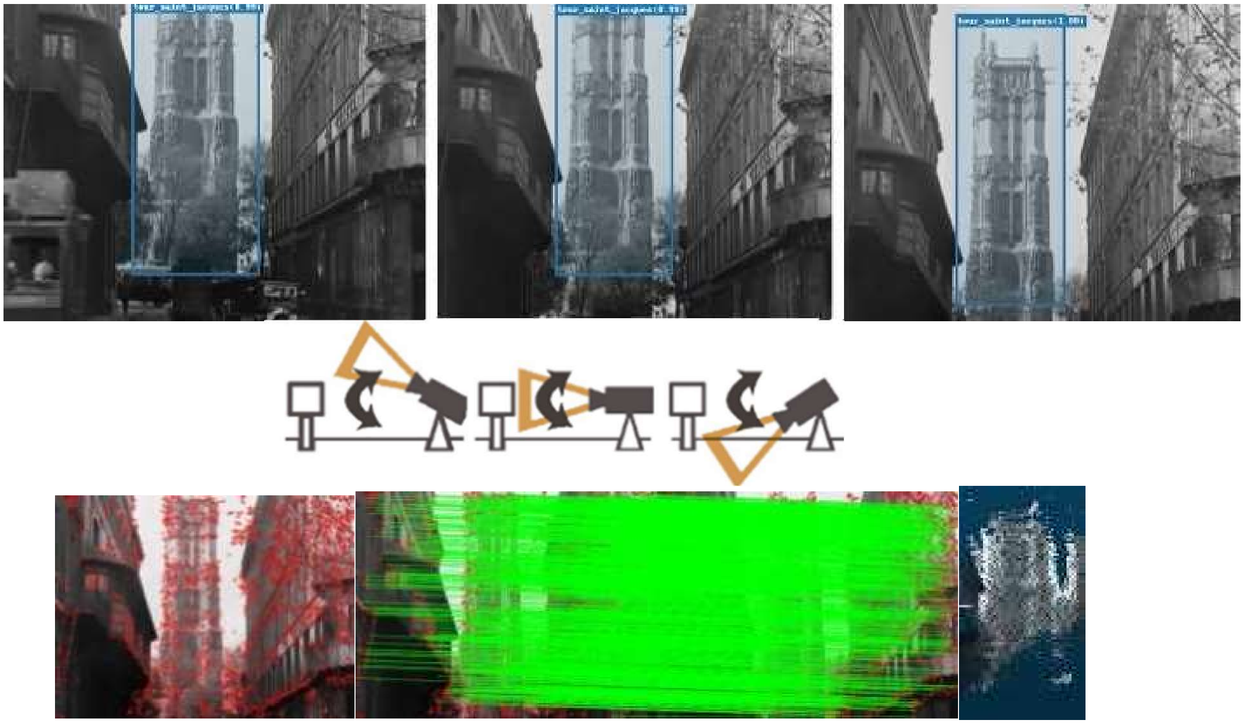


Figure 78. The frames identified from the film footage "Études sur Paris" with the object detection Neural Networks and the camera motion algorithm are selected and used for the photogrammetric process. The selection of the feature points, the results of the matching process and the final point clouds are obtained.

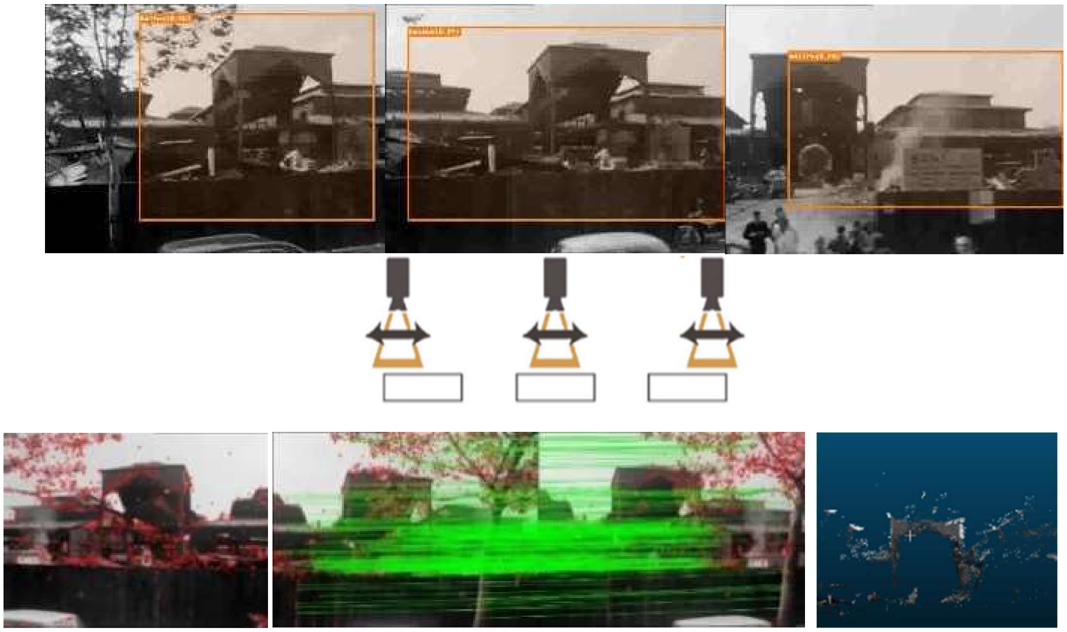


Figure 79. The frames identified from the film footage “La Destruction des Halles de Paris” with the object detection Neural Networks and the camera motion algorithm are selected and used for the photogrammetric process. The selection of the feature points, the results of the matching process and the final point clouds are obtained.

Metric quality assessment – Benchmark comparison

In order to assess the quality of the point clouds obtained from the photogrammetric process, the pipeline reported in the flowchart in Figure 80 is implemented.

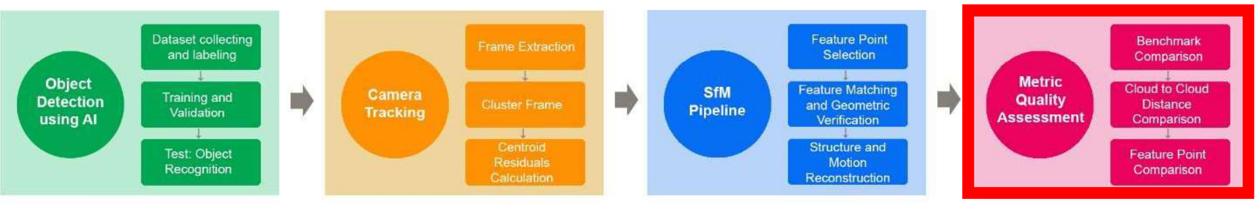

Object Detection using AI
Dataset collecting and labeling
Training and Validation
Test: Object Recognition
Camera Tracking
Frame Extraction
Cluster Frame
Centroid Residuals Calculation
SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction
Metric Quality Assessment
Benchmark Comparison
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 80. The general workflow proposed in this dissertation in which the fourth step concerning the metric quality assessment is highlighted in red and reported below.

Metric Quality Assessment
Benchmark Comparison
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 81. A focus on the fourth step of the general workflow proposed in this dissertation in which the three steps of the metric quality assessment are reported. In particular the first stage of the analysis concerns the benchmark comparison and it is highlighted in blue.

First of all, both the point clouds of the Tour Saint Jacques and Les Halles were compared with the benchmark reported in Chapter 2.

In order to do that, the values of the Residuals, resulted from the report of the photogrammetric process, were used for the estimation of the Mean and Standard Deviation. In addition, the Minimum and Maximum values were converted to centimetres with the calculation of the Ground Sample Distance (GSD), considering for the transformation the point closest to the camera.

The results are set out in Table 10 for the Tour Saint Jacques, in which a distance of 38 m was considered. Consequently the reference GSD for the tilting benchmark is 3.56 [cm/px] as shown in Table 9, and the GSD for the case of the tower is 5.06 [cm/px], both calculated on the point closest to the camera. Moreover, the graph in Figure 83 analyses the trend of the data between the benchmark and the Tour Saint Jacques case study.

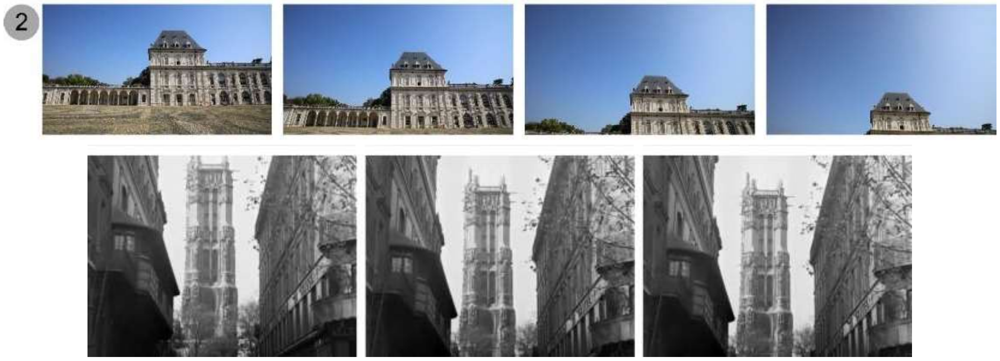


Figure 82. Frames extracted from the video sequences 2 of the benchmark of the left wing of the courtyard of the Valentino Castle shot with tilting camera motion at a distance of 38 m, compared with the frame extracted from the video “Études sur Paris”.

Table 9. Values of the benchmark, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding tilting camera motion case. The Tilting2 case, chosen as a reference in this processing concerning the Tour Saint Jacques for the comparison with the results of the frame extracted from the video "Études sur Paris", is highlighted in red.

Camera motion	Mean	Standard Deviation	Min Residual	Max Residual	Distance	GSD	Mean	Standard Deviation	Min Residual	Max Residual
	[px]	[px]	[px]	[px]	[m]	[cm/px]	[cm]	[cm]	[cm]	[cm]
Tilting1	0.4	0.1	0.1	0.6	43.0	4.0	1.5	0.5	0.5	2.3
Tilting2	0.5	0.2	0.1	0.7	38.0	3.6	1.7	0.6	0.4	2.5
Tilting3	0.5	0.1	0.1	0.7	33.0	3.1	1.4	0.4	0.4	2.0
Tilting4	0.5	0.1	0.1	0.7	28.0	2.6	1.3	0.4	0.3	1.8
Tilting5	0.5	0.1	0.1	0.9	18.0	1.7	0.9	0.2	0.2	1.5
Tilting6	0.5	0.1	0.1	0.7	10.0	0.9	0.5	0.1	0.1	0.7
Tilting7	0.6	0.1	0.2	0.7	5.0	0.5	0.3	0.1	0.1	0.3

Table 10. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals of the results of the photogrammetric processing of the frame extracted from the video "Études sur Paris", compared with the Tilting2 case of the benchmark.

Case	Mean	Standard Deviation	Min Residual	Max Residual	Distance	GSD	Mean	Standard Deviation	Min Residual	Max Residual
	[px]	[px]	[px]	[px]	[m]	[cm/px]	[cm]	[cm]	[cm]	[cm]
Benchmark	0.5	0.2	0.1	0.7	38.0	3.6	1.7	0.6	0.4	2.5
Case study	0.2	0.1	0.1	0.4	38.0	5.1	1.2	0.3	0.5	1.8

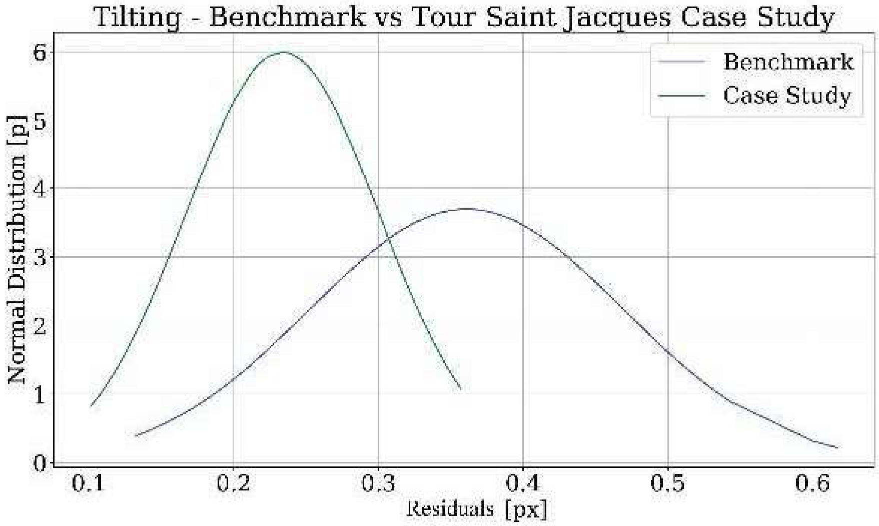


Figure 83. Comparison of Normal Distribution of the Residual value between benchmark tilting2 case and case study of the Tour Saint Jacques.

The graph shows that the curves in both the case study and the benchmark follow the Gaussian distribution. What it is noted from Table 10 is that was not find a significant difference in terms of Residuals values when comparing the two results. However, some little disparities are present. A possible explanation for this might be the approximation about the focal length of the camera used to shoot the film footage and the taking distance that could generate the observed discrepancies. For this reason, a margin of error has to be considered in this evaluation.

The same evaluation was performed for the case of Les Halles, as summarised in Table 12. A distance of 45 m was considered, and consequently the GSD for trucking benchmark is 4.22 [cm/px], and the GSD calculated for Les Halles is 9.14 [cm/px], both calculated on the point closest to the camera.

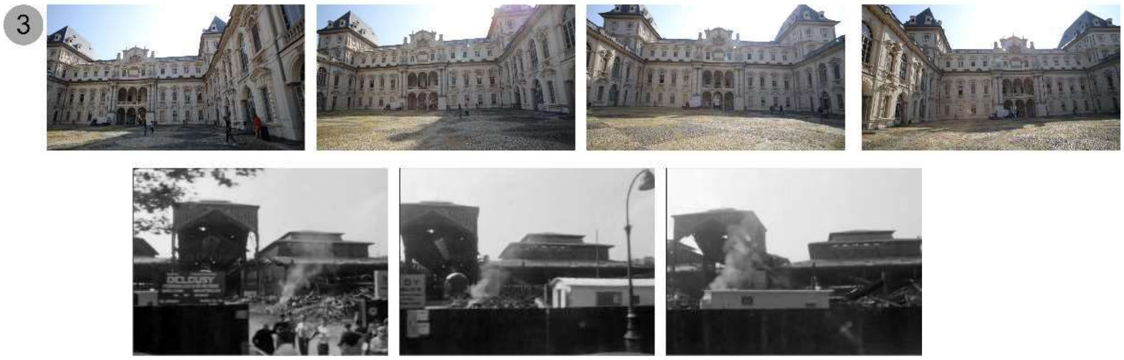


Figure 84. Frames extracted from the video sequences 3 of the benchmark of the façade of the Valentino Castle shot with trucking camera motion at a distance of 45 m, compared with the frame extracted from the video “La Destruction des Halles de Paris”.

Table 11. Values of the benchmark, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding trucking camera motion case. The Tilting3 case, chosen as a reference in this processing concerning Les Halles for the comparison with the results of the frame extracted from the video “La Destruction des Halles de Paris”, is highlighted in red.

Camera motion	**Mean**	**Standard Deviation**	**Min Residual**	**Max Residual**	**Distance**	**GSD**	**Mean**	**Standard Deviation**	**Min Residual**	**Max Residual**
	[px]	**[px]**	**[px]**	**[px]**	**[m]**	**[cm/px]**	**[cm]**	**[cm]**	**[cm]**	**[cm]**
Trucking1	0.6	0.1	0.3	0.9	85.0	8.0	5.0	0.9	2.3	7.5
Trucking2	0.7	0.1	0.2	0.8	65.0	6.1	3.9	0.5	1.4	4.7
Trucking3	0.7	0.1	0.3	0.8	45.0	4.2	2.9	0.4	1.2	3.4
Trucking4	0.7	0.1	0.2	1.1	25.0	2.3	1.6	0.3	0.6	2.6

Table 12. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals of the results of the photogrammetric processing of the frame extracted from the video "La Destruction des Halles de Paris", compared with the Trucking3 case of the benchmark.

Case	Mean	Standard Deviation	Min Residual	Max Residual	Distance	GSD	Mean	Standard Deviation	Min Residual	Max Residual
	[px]	[px]	[px]	[px]	[m]	[cm/px]	[cm]	[cm]	[cm]	[cm]
Benchmark	0.7	0.1	0.3	0.8	45.0	4.2	2.9	0.4	1.2	3.4
Case study	0.5	0.1	0.2	0.8	45.0	9.1	4.6	0.9	1.6	7.2

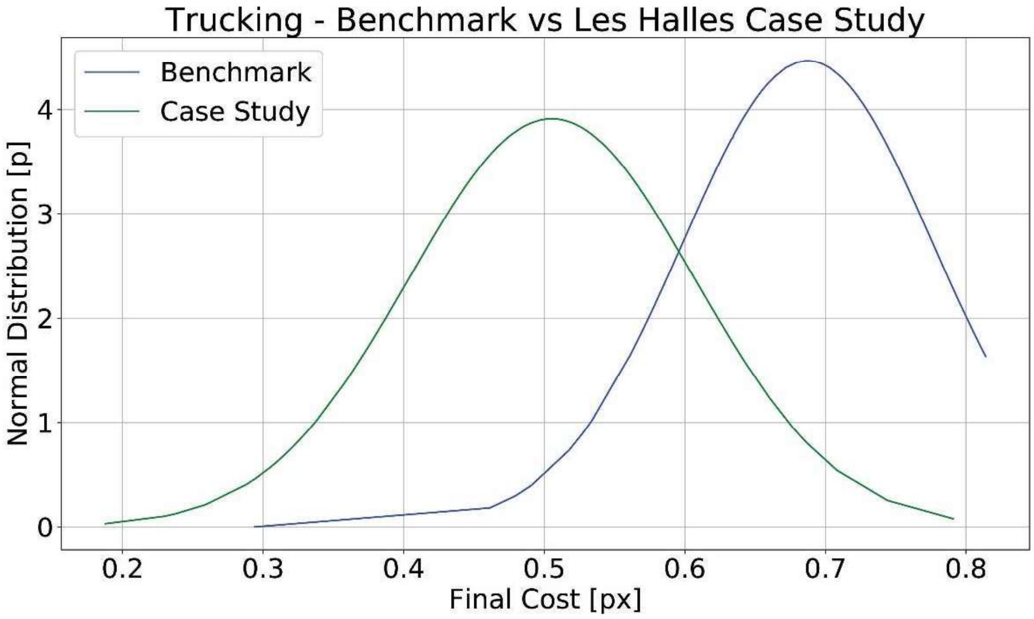


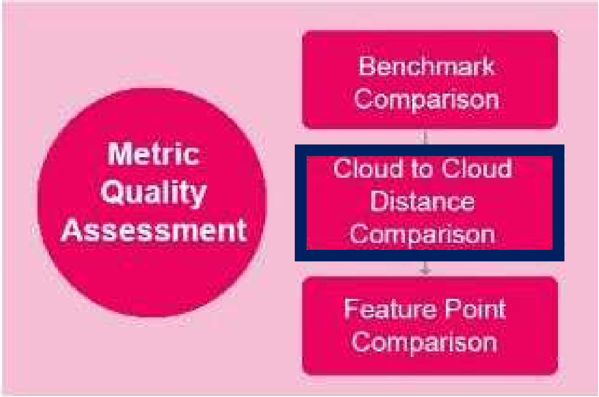
Figure 85. Comparison of Normal Distribution of the Residual value between benchmark trucking3 case and case study of Les Halles.

Even in this case, both the curves of the benchmark and the case study follows the Gaussian distribution. The observed differences in terms of Residuals values between the pavilion and the benchmark are not significant, even if the case study presents values a little bit greater than the benchmark.

Anyway, the comparison of the two results reveals a high-quality precision of the photogrammetric reconstruction, in both cases almost at the level of the benchmark.

Metric quality assessment – Cloud to mesh distance comparison

Once analysed the precision of the 3D model obtained from the photogrammetric process, it is useful to evaluate the accuracy of them. In order to do that, the next step of the metric quality evaluation was concerned with the scale and the analysis of the point cloud obtained from the photogrammetric process. This was performed in two different ways: the cloud to cloud (mesh in this case) distance comparison and the feature point comparison (Figure 85).



Metric Quality Assessment

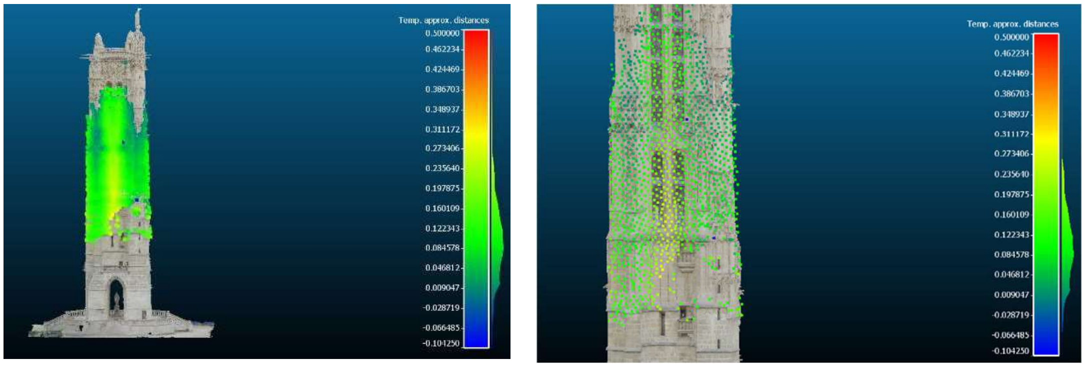
Benchmark Comparison

Cloud to Cloud Distance Comparison

Feature Point Comparison

Figure 86. A focus on the fourth step of the general workflow proposed in this dissertation in which the three steps of the metric quality assessment are reported. In particular the second stage of the analysis concerns the cloud to cloud (mesh) distance comparison and it is highlighted in blue.

In the case of the Tour Saint Jacques, the point cloud obtained by the photogrammetric process, although of low density, was compared with the 3D model of Iconem. The comparison (Figure 87) showed that the computed distances between the model mesh and the resulting point cloud were less than 0.5 pixels.



(a) (b)

Figure 87. Cloud to mesh distance comparison between the point cloud obtained from the photogrammetric processing of the video “Études sur Paris” and the 3D model obtained from the 3D survey by Iconem. (a) View of the entire Tour Saint Jacques. (b) A detail of the comparison.

Metric quality assessment – Feature point comparison

The specific Feature Points corresponding to points of interest in project drawing manually selected during the photogrammetric process, were used in the final part of the metric evaluation (see Chapter 2).

Metric Quality Assessment
Benchmark Comparison
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 88. A focus on the fourth step of the general workflow proposed in this dissertation in which the three steps of the metric quality assessment are reported. In particular the third stage of the analysis concerns the Feature Point comparison and it is highlighted in blue.

For the case of the tower this was useful to compare the survey from the architect Gabriel Davioud's drawing before the Haussmann transformation (Figure 89), with the 3D model of the recent survey from Iconem, chosen as reference, and the point cloud obtained from the photogrammetric process of the film footage. The results of the comparison are reported in Table 13.

Table 13. Residuals values resulted from the comparison of the distance between the 3D model by Iconem, the point cloud resulted from the photogrammetric processing of the video "Études sur Paris" and the drawing by the architect Gabriel Davioud.

Distances	Iconem [m]	Point Cloud [m]		Drawing [m]	
		Values	Residuals	Values	Residuals
AB	1.60			1.88	-0.28
BC	7.28			7.53	-0.25
DE	6.00			7.62	-1.62
FG	6.50	6.40	0.10	9.48	-2.98
HI	15.76	15.50	0.36	19.68	-3.92
IL	16.70	16.70	0.10	27.50	-10.8
MN	10.00	10.0	0.00	17.17	-7.17
OP	0.70	0.70	0.00	1.50	-0.8
QR	7.00	7.00	0.00	11.46	-4.46
ST	1.50	1.40	0.10	2.00	-0.5

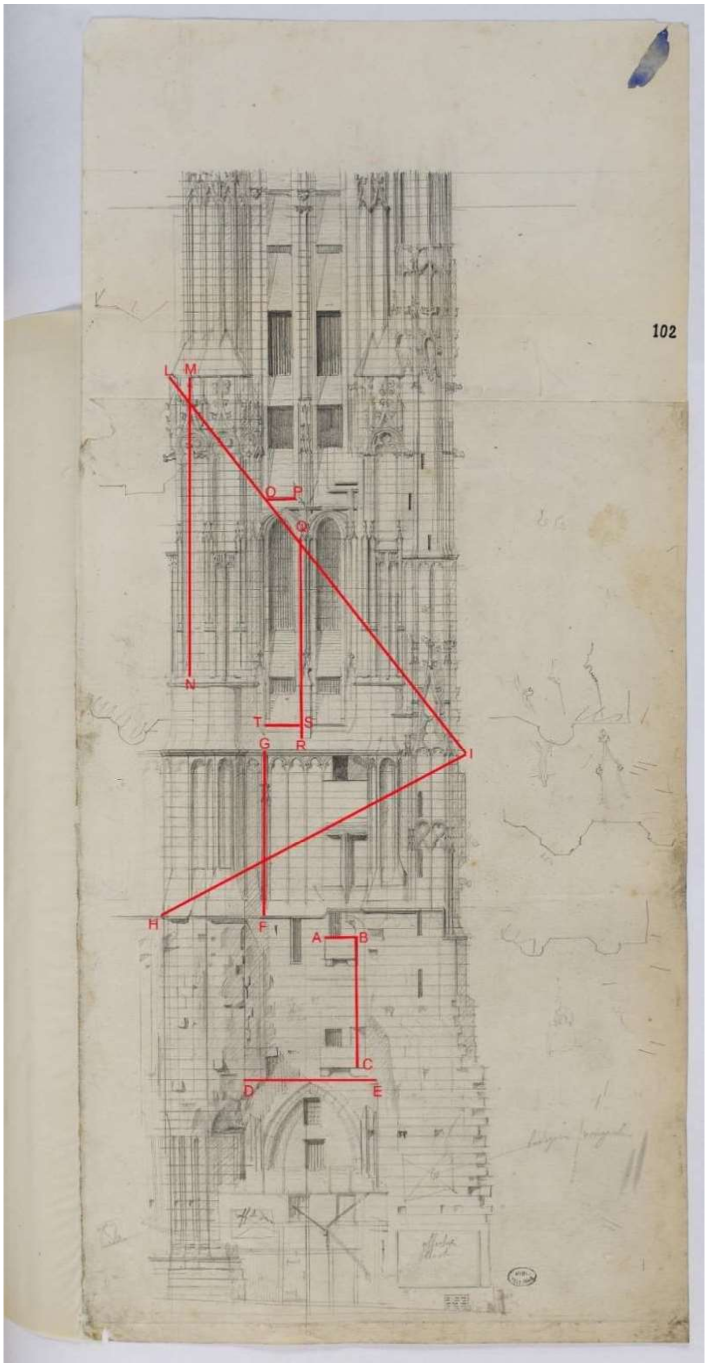


Figure 89. Distances extracted from the drawing by the architect Gabriel Davioud of the Tour Saint Jacques.

As already noted in the cloud to mesh comparison, the differences between the two point clouds are not significant, also in terms of Residuals, as shown in Table 13, instead of the survey by the architect Gabriel Daviuod that presents high differences with the real measures. It has to be considered that the drawing was made before the transformation and by hand, therefore the differences especially in the vertical measures are according to the expectations.

The method of Feature Point extraction was particularly useful for the case of Les Halles because the only 3D models available are those of the part of the pavilion that survived the destruction, a 3D model of the upper part of the pavilion made by the Commune de Nogent-sur-Marne (Figure 92) and a photogrammetric point cloud of the lower part of the pavilion moved in Yokohama, Japan, made on site (Figure 91b). These models refer to the pavilion n°8 and certainly during the reconstruction some differences from the original building occurred. Moreover, the metric description of some parts of the pavilions was reported in Lemoine, 1980. Nevertheless, they could be used, together with the architect Victor Baltard's project drawings (Figure 90) to compare some distances extracted from the obtained point cloud. Thanks to the presence of these feature points in the final point clouds (Figure 91a), it was possible to scale the obtained photogrammetric model using the distance AG because this part of the pavilion is well visible in the point cloud. These points are then used to extract some distances from all the sources available and the results are reported in Table 14 and 15.

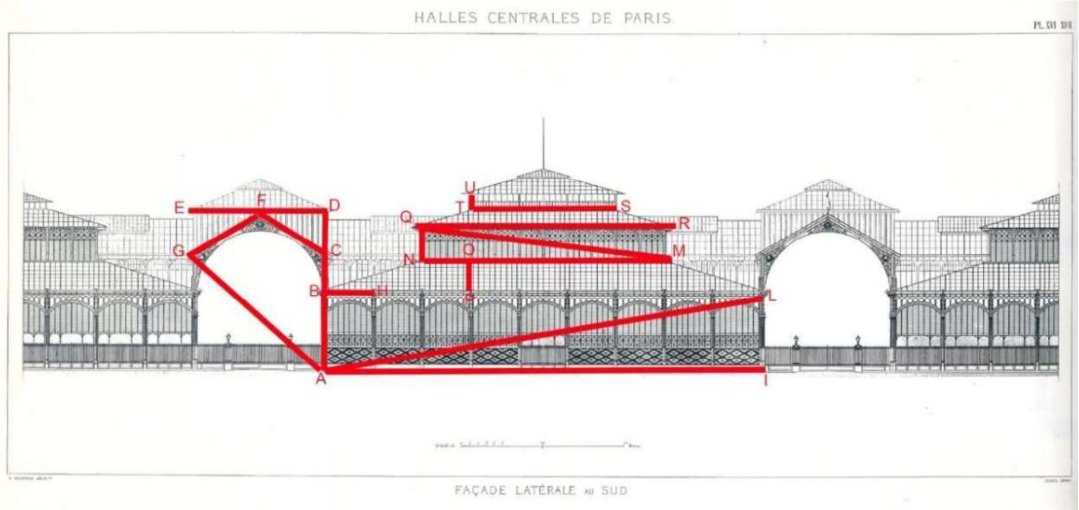
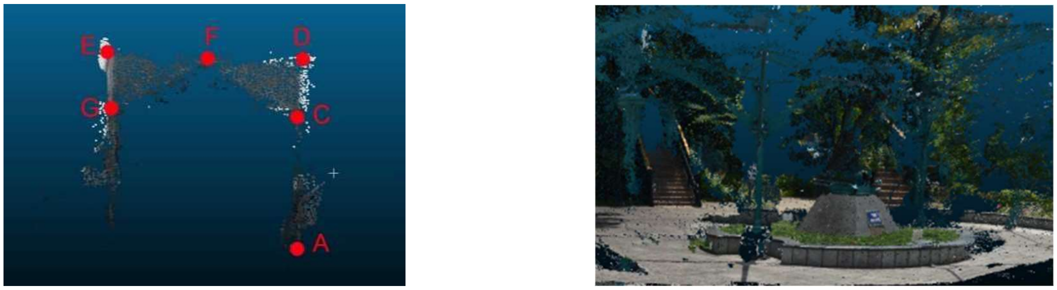


Figure 90. Distances extracted from the drawing of Façade of Les Halles, Victor Baltard, Monographie des Halles centrales, 1863, Archives de Paris.



(a) (b)

Figure 91. Point clouds of the Pavilion No. 8: (a) the point cloud has resulted from the photogrammetric processing of the film "La Destruction des Halles de Paris". (b) the point cloud has resulted from a photogrammetric survey of the pavilion's cave in Yokohama (Japan).

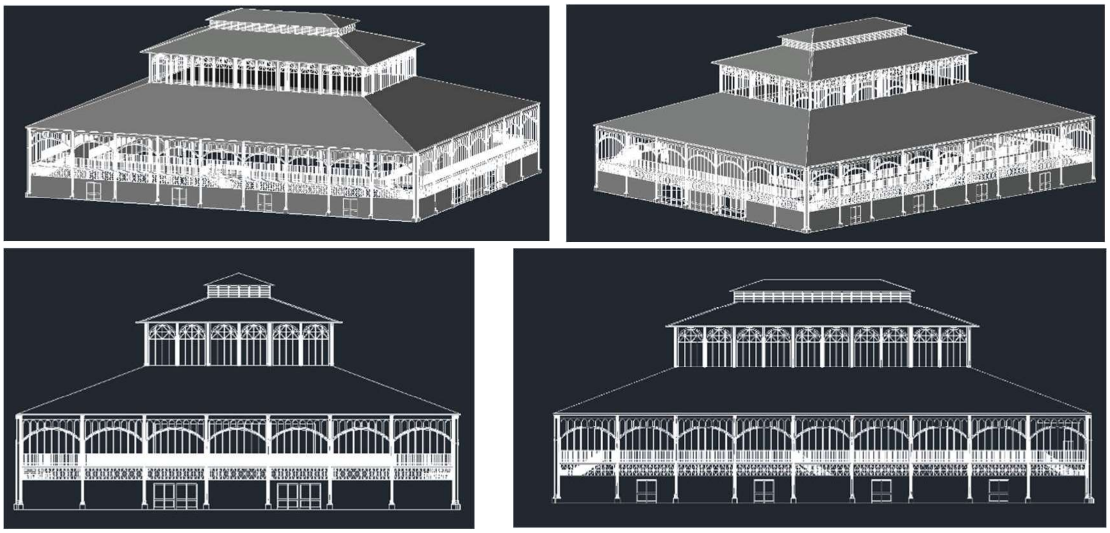


Figure 92. 3D model of the upper part of the Pavilion No. 8 made by the Commune de Nogent-sur-Marne, 2019.

Table 14. Distances values measured in project drawing of the architect Victor Baltard, the 3D model by the Commune de Nogent-sur-Marne, the description in Lemoine, 1980 and the point cloud resulted from the photogrammetric process of the video "La Destruction des Halles de Paris".

Distance	**Project Drawings**	**3D model**	**Description in Lemoine, 1980**	**Point Cloud**
	[m]	**[m]**	**[m]**	**[m]**
AB	8.50	9.00		
BC	4.25			
AC	11.55	13.60	12.50	
CD	4.60			4.59
CF	9.50			9.58
DE	15.50			15.79
FG	9.50			9.58
AG	18.54			18.87
BH	5.00	6.00		
AI	47.50	55.00		
AL	47.66	63.37		
MN	27.76	30.00		
OP	3.05	4.60		
NQ	3.80	4.00		
AB+OP+NQ	15.35	17.60	18.00	
QR	28.43	32.00		
MQ	26.54	29.93		
ST	15.80	18.00		
TU	1.70	1.20		

Table 15. Distances values measured in project drawing of the architect Victor Baltard, the description in Lemoine, 1980 and the point cloud resulted from the photogrammetric survey of the pavilion's cave in Yokohama (Japan).

Distance	Project Drawings	Description in Lemoine, 1980	Point Cloud
	[m]	[m]	[m]
AB	4.00		4.00
BC	4.50		5.10
CD	1.15		1.15
AE	0.55	0.44	0.47
AF	10.00	12.00	12.00

Despite certain limitations due to the lack of information on the technical features of the camera and the film used to shoot the videos and the unavailability of a precise 3D model for the case study of Les Halles, since it no longer exists, these findings are very encouraging for the metric certification of the models obtained.

With this last evaluation, the implementation of the entire workflow on historical film footage is completed.

4.3 Validation of the photogrammetric method

In order to validate the workflow proposed in this dissertation, further case studies were chosen. The validation was implemented only on the photogrammetric method to show its potentialities. For this reason, it was applied to different situation and asset of the heritage. First, it was applied to historical photographs with the aim to reconstruct and compare heritage that has been restored. Then the same method was applied to historical videos reconstructing ephemeral architecture such as the pavilions of International Exposition.

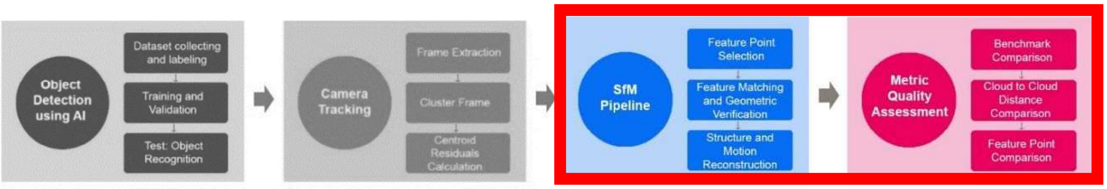

Object Detection using AI
Dataset collecting and labeling
Training and Validation
Test: Object Recognition
Camera Tracking
Frame Extraction
Cluster Frame
Centroid Residuals Calculation
SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction
Metric Quality Assessment
Benchmark Comparison
Cloud to Cloud Distance Comparison
Feature Point Comparison

Figure 93. The general workflow proposed in this dissertation: for the two case studies analysed the validation was performed only in the last two steps: the SfM pipeline and the Metric quality assessment.

4.3.1 Photogrammetric reconstruction from historical photographs of restored Cultural Heritage

First of all, the method was tested on the case of restored Cultural Heritage, selecting two different scenarios. This research was conducted during the period abroad at the Tokyo Institute of Technology and in collaboration with Prof. Nasu, Prof. Higuchi and Prof. Sugawara, that provided the material for the analysis of two case studies.

The first case study was to examine the historical and architectural transformations of the church of Karanlık in Göreme National Park (Cappadocia, Turkey) by comparing the reconstruction from historical archive photographs (by G. de Jerphanion, 1925-42) and the present state of the church with special attention to the wall paintings. The second was to measure the difference between the building of the former Matsuno-Yu (Public Bath) in "Naka-machi Komise Street" (Japan) before and after the restoration by comparing the point cloud obtained from images from local archives (2013) and a recent survey of the building (2018).

Churches in Göreme National Park and the Rock Sites of Cappadocia – Restored wall paintings

“The Göreme National Park and the Rock Sites of Cappadocia, registered UNESCO’s World Heritage list in 1985, is located in Nevşehir Province, 280 km southeast of Ankara. It is in this spectacular landscape, wherein rock-cut churches provide unique evidence of Byzantine art from the post-iconoclastic period, especially during the 9th and 10th centuries (criteria for selection (i)).

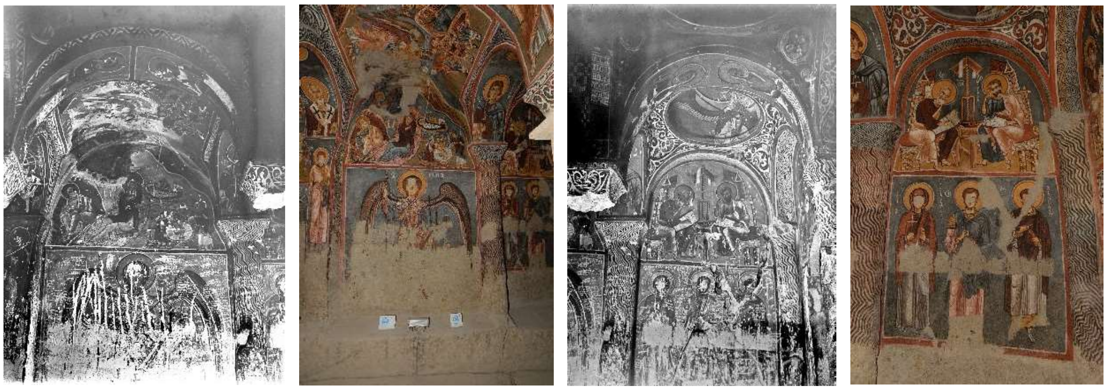


Figure 94. Historical archives photographs from G. de Jerphanion’s work (1925-42) and images of the actual state of the paintings in Karanlık Kilise (11th century).

The valley marks the centre of Cappadocia and has more than 100 churches from the end of the 9th century to the 11th century. It was only after the publication of G. de Jerphanion’s work (1925-42) that the importance of the Cappadocian Churches was put into evidence. The French Jesuit professor had explored the eastern part of Cappadocia and left great documentation of the rock-cut churches with many drawings and photographs. But the condition of the rock-cut churches is not the same as Jerphanion's period. The spectacular landscape of Cappadocia depends on the fact that most of the rock here is tuff, which is easily eroded. In other words, the rocks from which the churches were hollowed out are gradually disappearing. And extended abandonment of the churches also brought along artificial damages, such as malicious scars and graffiti. In addition, the rapidly increasing number of tourists harms the environment of the preservation of the monuments" (Higuchi et al., 2019). One of the famous masterpieces of this area is the Karanlık Kilise (11th century) and it was chosen as a case study (Figure 94).

Nakamachi Komise Street in Japan – Restored architecture

Kuroishi is a city which locates on the northern edge of Honshu, Japan's largest main island. The city has a traditional landscape from the Japanese feudal period (Edo period, 17th century). Here, every traditional house has a wooden arcade, Komise, in front of the building, and the arcades continue uninterrupted (Shimazu and Nasu, 2019; Kuroishi, 2005). Komise protects pedestrians from snow, rain and sunlight in summer. Due to its characteristic urban landscape, the central part of the city, Naka-machi, was designated as an important protected area for traditional groups of buildings in 2005. Because Naka-machi Street is so attractive, it is also called one of the 100 largest streets in Japan. The case study analysed is a building along this street with Komise, the former Matsuno-Yu building. It was a traditional Japanese public bath, but it was closed in 1993. After Naka-machi became an important conservation area for traditional building groups, the building became a community centre in 2015 with the support of the residents. For the processing, pictures before restoration were taken by locals in 2013 (Figure 95) and pictures after restoration were taken by prof. Nasu and his students during their on-site survey in 2018 (Figure 96).

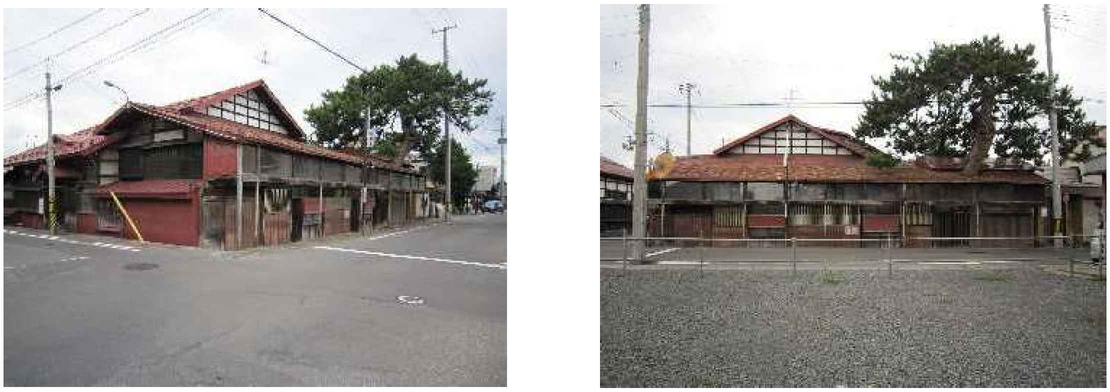


Figure 95. Local archive photographs of the Former Matsuno-Yu building before the restoration in 2013.

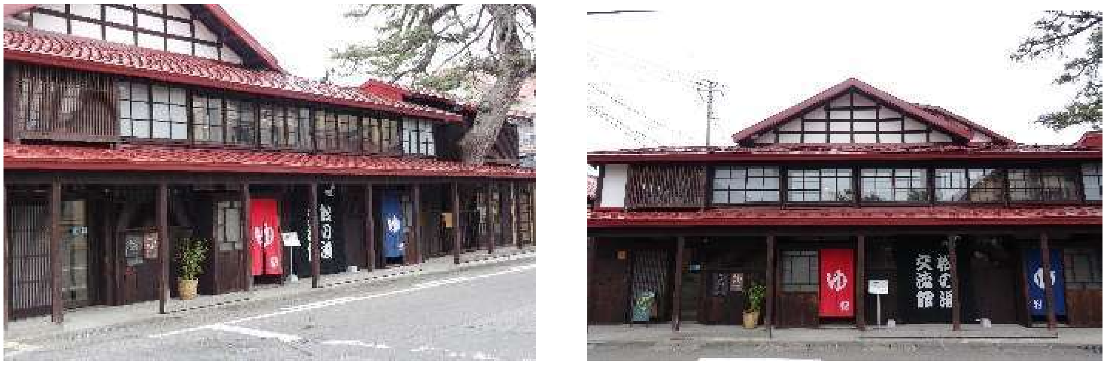


Figure 96. Images of the photographic on-site survey of the Former Matsuno-Yu building after the restoration in 2013, prof. Nasu, 2018.

Results of the application of the method to the two case studies

The results of the last two steps of the proposed workflow implementation (Figure 96) for the two case studies are then discussed.

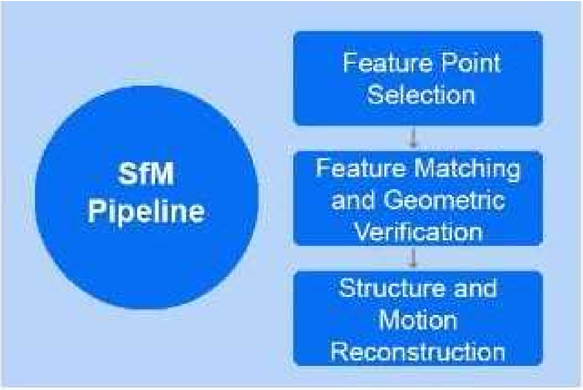

SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction

Figure 97. A focus on the third step of the general workflow proposed in this dissertation in which the three steps of the photogrammetric pipeline are reported.

With the standard “Feature detection and extraction” step some keypoints were recognized, but, as shown in Figure 98, some important radiometric corners in the photograph, which also appear in other images, are missing.

With the “Feature point selection” step, introduced in this research, it is possible to manually detect the feature point of interest. The coordinates of the searched point were measured with the WebPlotDigitizer tool (Figure 99) (https://automeris.io/WebPlotDigitizer, October 2019) and inserted in the software with the two methods previously described in Chapter 2.

This manual selection was performed for all the feature points chosen in the archive images for both case studies (as shown in Figure 100 and 101). Then, they were used in the matching process and appear in the final point cloud.

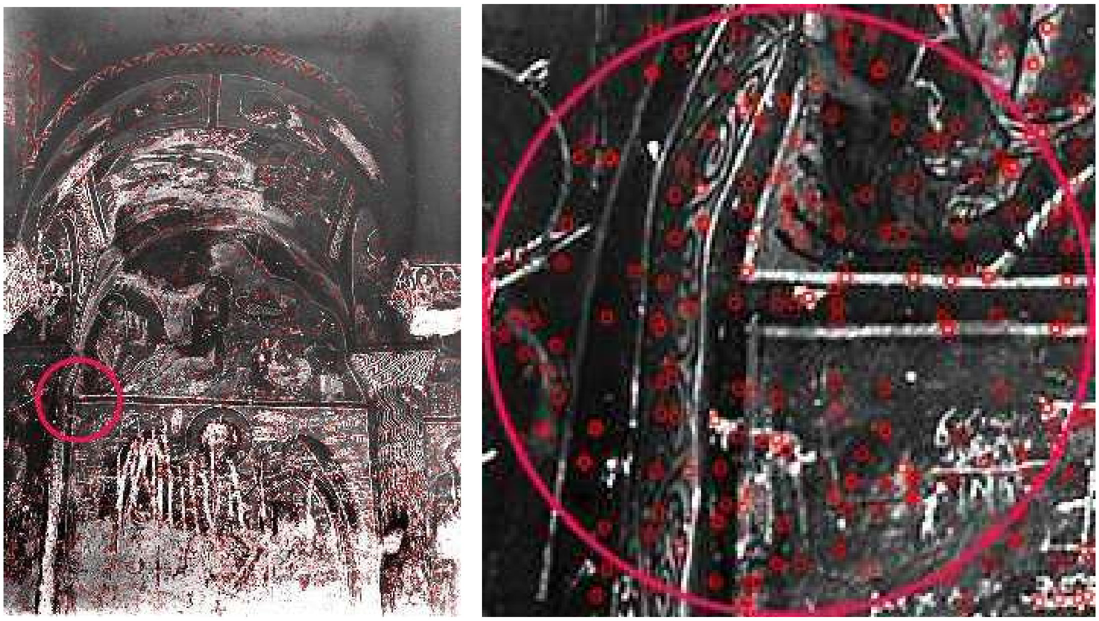
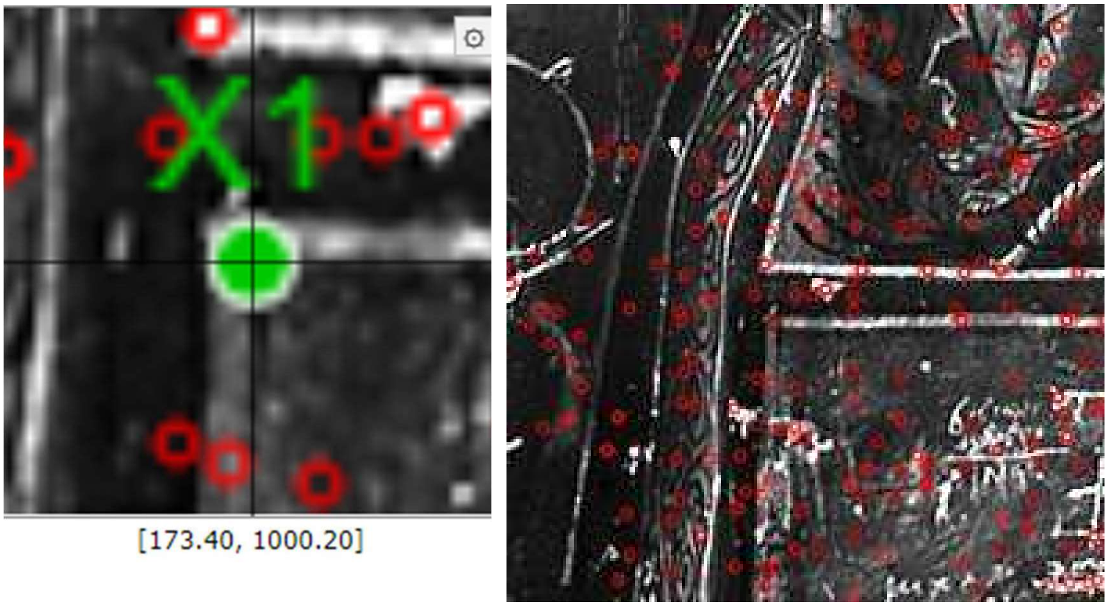


Figure 98. Keypoints automatically extracted from the software with the standard "Feature detection and extraction" step: some important radiometric corners in the photograph are missing.



[173.40, 1000.20]

Figure 99. Image coordinates of feature point measured with WebPlotDigitizer tool and added in COLMAP with the "Feature point selection" step.

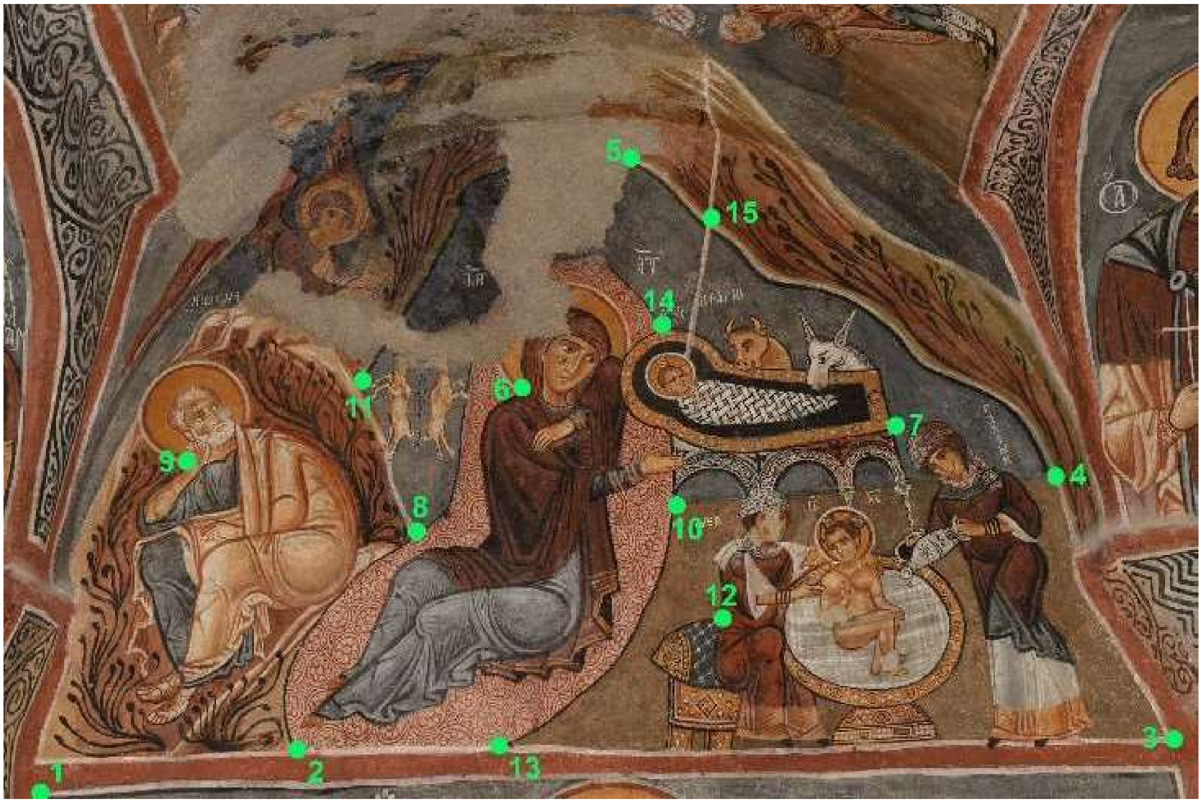


Figure 100. Feature points manually chosen for the Karanlık church.

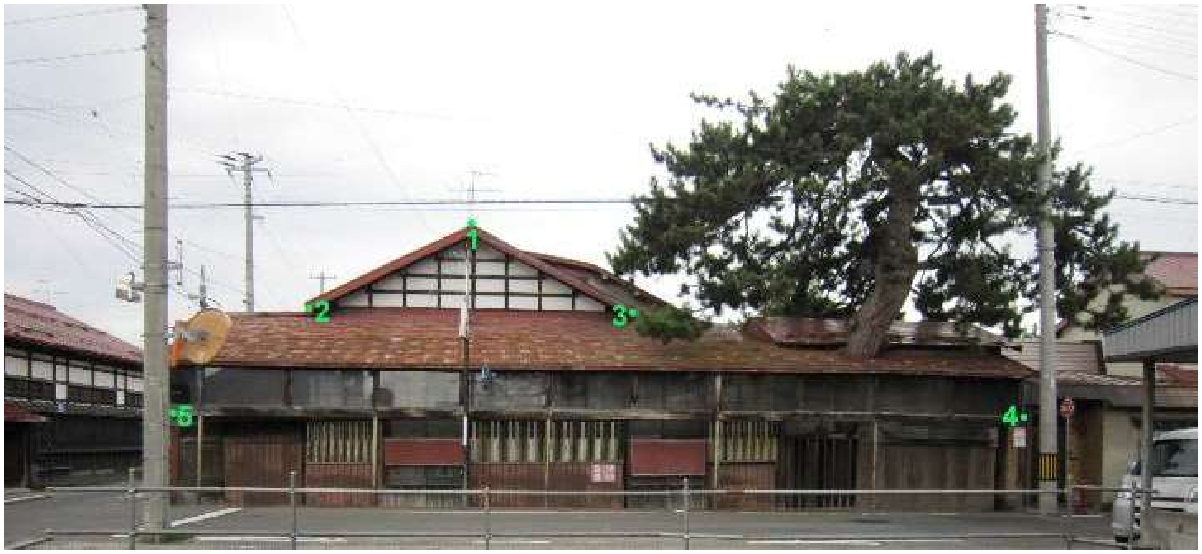
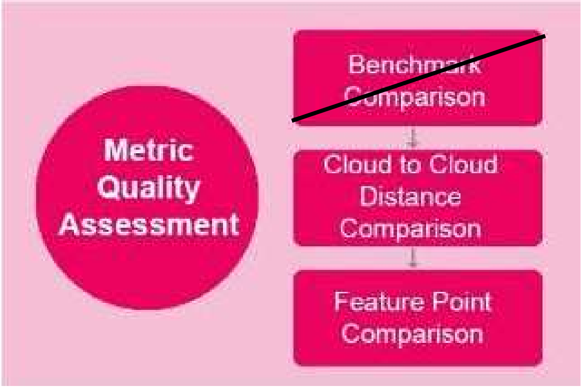


Figure 101. Feature points manually chosen for the Former Matsuno-Yu building.

Thanks to the presence of these points, it was possible to perform the distance comparison in CloudCompare (Figure 102) between the sparse point clouds obtained from the photogrammetric processing of archival photographs and the point clouds obtained from the photogrammetric recent survey for both case studies.



Metric Quality Assessment

~~Benchmark Comparison~~

Cloud to Cloud Distance Comparison

Feature Point Comparison

Figure 102. A focus on the fourth step of the general workflow proposed in this dissertation in which the three steps of the metric quality assessment are reported.

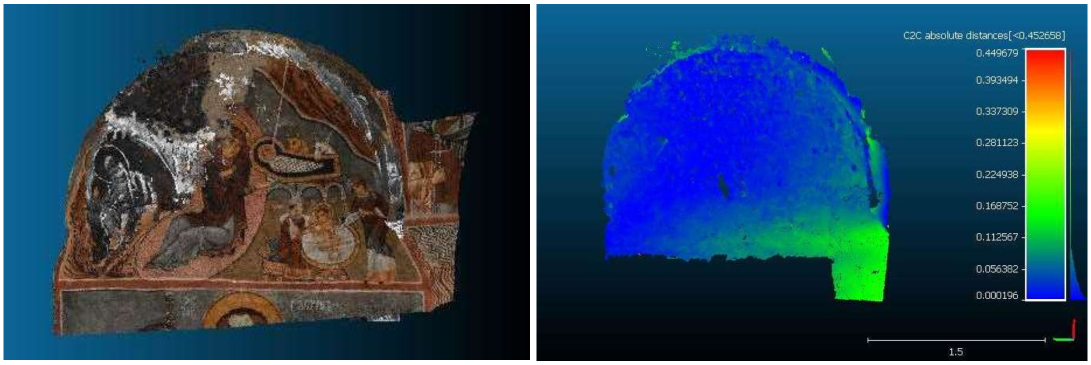


Figure 103. Cloud to cloud distance for the Karanlık church in which the point cloud obtained from the photogrammetric reconstruction from historical images is compared with the point cloud obtained from a photogrammetric recent survey of the church.

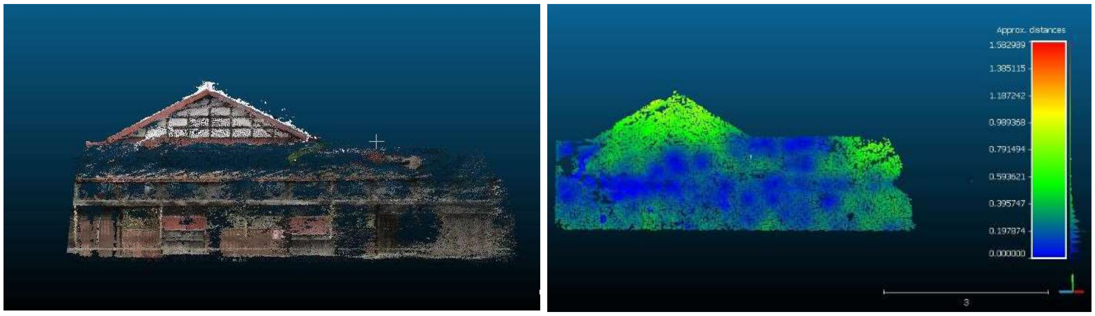


Figure 104. Cloud to cloud distance for the Former Matsuno-Yu building in which the point cloud obtained from the photogrammetric reconstruction from archival images is compared with the point cloud obtained from a photogrammetric recent survey of the building.

Figure 103 shows that the distance between the two point clouds is less than 0.1m for the first case study. For the second case study, the distance between the two point clouds (Figure 104) presents a greater value, until 1 m. It is a demonstration that the restoration works widely modified the shape of the building.

For this reason, it was chosen to estimate the Residual values only for the first case study. In the following Table 16 the results of the feature point comparison are reported.

As shown, the Mean of the Residuals estimated on the three coordinates is summed it up:

- Mean ΔX = - 0,04 m; Standard Deviation = 0,06 m
- Mean ΔY = 0,04 m; Standard Deviation = 0,08 m
- Mean ΔZ =0,03 m; Standard Deviation = 0,04 m.

Table 16. Residuals values (ΔX , ΔY and ΔZ) of the feature points measured on the point cloud obtained from the processing of historical photographs (called X_{old}, Y_{old} and Z_{old}) and the point cloud of the recent photogrammetric survey (called X_{new}, Y_{new}, Z_{new}).

#	X_{new}	Y_{new}	Z_{new}	X_{old}	Y_{old}	Z_{old}	ΔX	ΔY	ΔZ
	[m]	[m]	[m]	[m]	[m]	[m]	[m]	[m]	[m]
1	0.46	1.77	-3.51	0.54	1.64	-3.60	-0.08	0.13	0.09
2	0.55	1.16	-3.61	0.61	1.04	-3.66	-0.06	0.13	0.06
3	1.17	0.80	-3.73	1.27	0.65	-3.80	-0.10	0.16	0.07
4	2.15	0.28	-3.76	2.20	0.39	-3.77	-0.05	-0.11	0.00
5	1.47	0.59	-3.69	1.50	0.59	-3.69	-0.03	0.00	-0.01
6	1.33	0.38	-3.72	1.40	0.27	-3.78	-0.07	0.11	0.05
7	1.08	0.86	-3.64	1.10	0.83	-3.64	-0.02	0.04	0.01
8	1.29	1.45	-3.58	1.30	1.38	-3.55	-0.01	0.07	-0.02
9	1.13	0.18	-3.69	1.19	0.20	-3.71	-0.06	-0.02	0.03
10	1.49	1.02	-3.66	1.35	0.90	-3.64	0.14	0.12	-0.02
11	0.82	0.09	-3.68	0.91	0.11	-3.74	-0.09	-0.02	0.06
12	0.53	0.65	-3.66	0.64	0.59	-3.74	-0.10	0.06	0.08
13	1.63	0.22	-3.73	1.60	0.25	-3.74	0.03	-0.02	0.01
14	1.97	0.06	-3.78	2.00	0.17	-3.80	-0.03	-0.11	0.02

4.3.2 Photogrammetric reconstruction from historical film footage of ephemeral architecture

Finally, the method was tested on the case of lost Cultural Heritage, selecting two different case studies in the scenario of ephemeral architecture of the International Exposition taken in Turin at the beginning of ‘900. Material was provided thanks to a collaboration with the I-Media-Cities platform in which the CINECA and the Cinema Museum of Turin are partners of the project (see Chapter 2).

The Hungarian pavilion at Turin International Exposition of 1911

The first case study chosen to validate the methodology proposed is the photogrammetric reconstruction of the temporary pavilion of Hungary (Figure 105). The pavilion was built for the Turin International Exposition of 1911 that took place in the park of Valentino.

In the context of baroque pavilions, usual in a baroque town as Turin, Hungary preferred to choose a national way to design its pavilion and got a great success. The amazing Hungarian pavilion designed by Emil Tòry and Mòric Pogàny in the official guide of the exhibition was the most important attraction. It was clear the goal of this special kind of architecture and decoration: to get a style both modern and national, melting Secession and popular decorative patterns. It was realized in wood, one of the most important Hungarian material, and presented geometrical lines and squared volumes (Cornaglia, 2001).

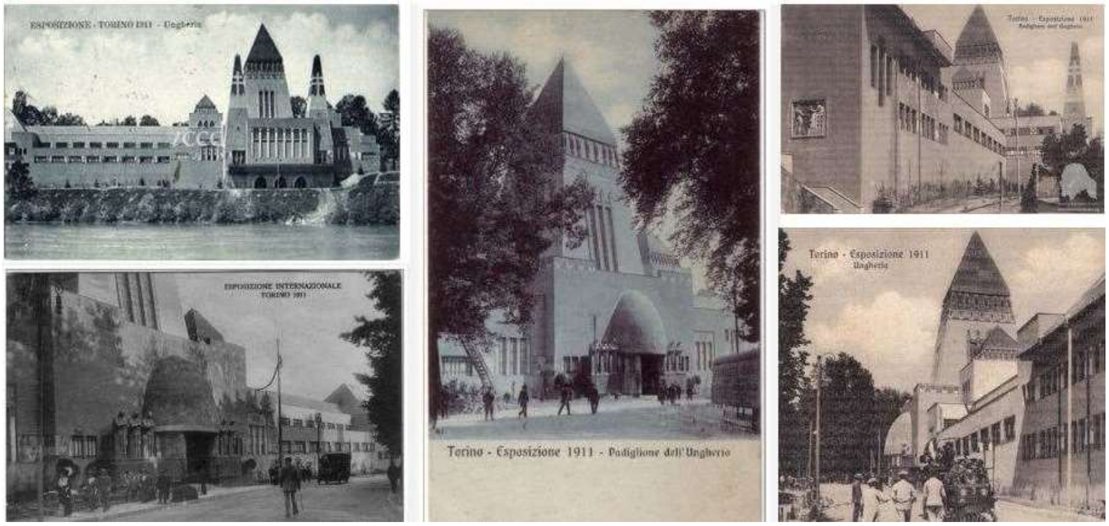

ESPOSIZIONE - TORINO 1911 - Ungheria
ESPOSIZIONE INTERNAZIONALE TORINO 1911
Torino - Esposizione 1911 - Padiglione dell'Ungheria
Torino - Esposizione 1911 Padiglione dell'Ungheria
Torino - Esposizione 1911 Ungheria

Figure 105. Pictures and postcards of the Hungarian pavilion in the Turin International Exposition of 1911.

The Mines and Ceramics pavilion at Turin International Exposition of 1928

The second case study is the Exposition of 1928 (May 1st - November 11th) held in Turin, in the park of Valentino, within which were built temporary pavilions of which now there are few traces.

The exhibition was part of a series of civil, military and economic events organised on the occasion of the 10th anniversary of the end of the First World War and the fourth centenary of the birth of king Emanuele Filiberto. In designing the pavilions, the architects had the opportunity to experiment and offer to the public new architectural languages in real buildings even if temporary.

By combining video, photo and design drawings (Figure 106-108), it was possible to extract metric information from the “Mines and Ceramics” pavilion by the architects G. Pagano-Pogatschnig and P. Perona.

The pavilion consisted of a two-level unit, high approximatively 16 m, and two parallel swings, of 5.50 m. It presents the characteristics of rigidly modern architecture: a triumph of the straight line, symmetrical harmony, equilibrium of the closed masses, rectangular windows aligned, corner doors, slight pillars.

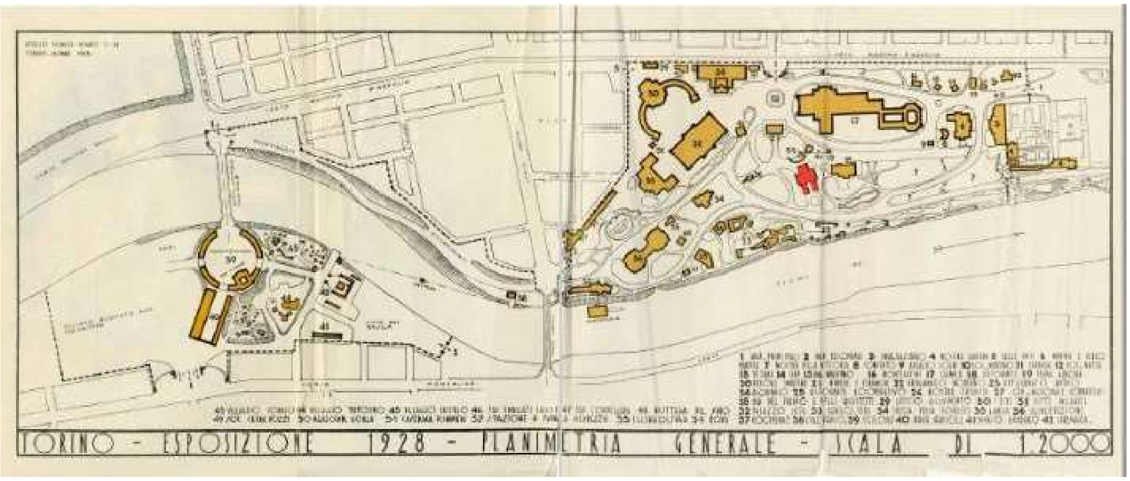

TORINO - ESPOSIZIONE 1928 - PLANIMETRIA GENERALE - SCALA DI 1:2000

Figure 106. Planimetry of Turin Exposition 1928 – In red the “Mines and Ceramics” pavilion, Library of Politecnico di Torino "Roberto Gabetti", 1928.

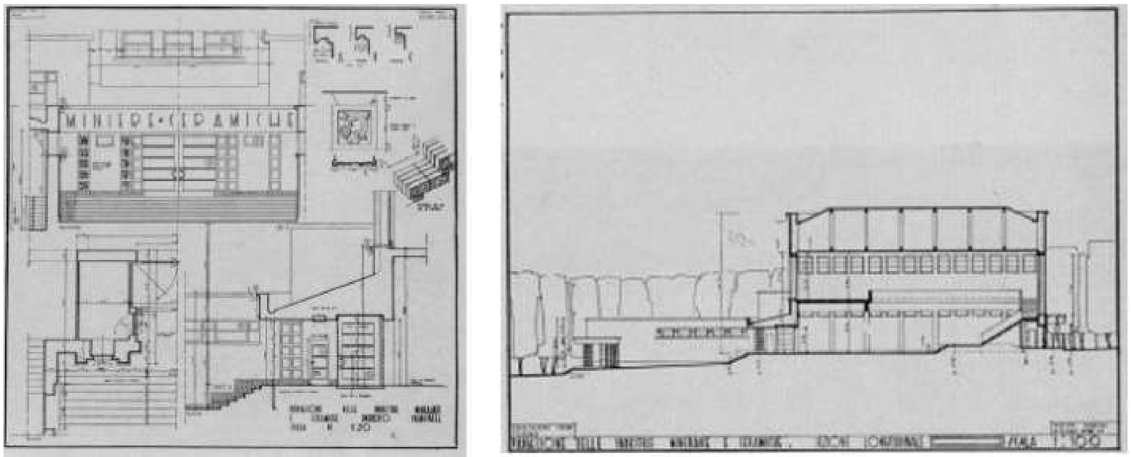


Figure 107. Project drawing of “Mines and Ceramics” pavilion at Turin Exposition 1928. Plan, Section and architectural details, Pagano et al., 1930.

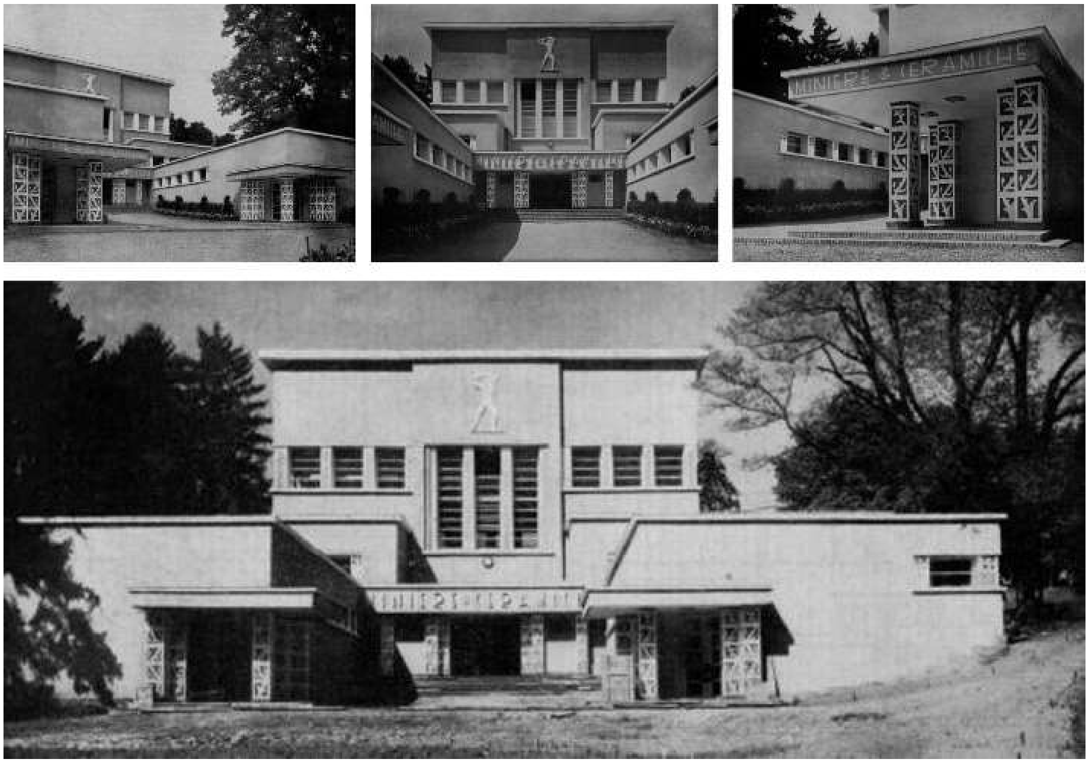


Figure 108. The “Mines and Ceramics” pavilion at Turin Exposition 1928, Pagano et al., 1930.

Photogrammetric reconstruction of the pavilions and metric quality assessment

As an ephemeral architecture, the pavilions were demolished, but, luckily a lot of documents such as project drawings photographs, postcards, watercolours, books and journals saved its historical memory. The Hungarian pavilion appeared in the historical film footage “Nei cantieri dell’esposizione” shot in 1911 (Figure 109), now stored in the Cinema Museum of Turin and available on I-Media-Cities platform. The film was shot with the Left/Right Motion or Trucking type of camera motion and presents the following characteristics in Table 17.

Table 17. Technical feature of the film footage “Nei cantieri dell’esposizione”.

Colour	Black & White
Gauge	35 mm (24x36 mm)
Focal length	25 mm
Digital format Resolution	720x540 pixels

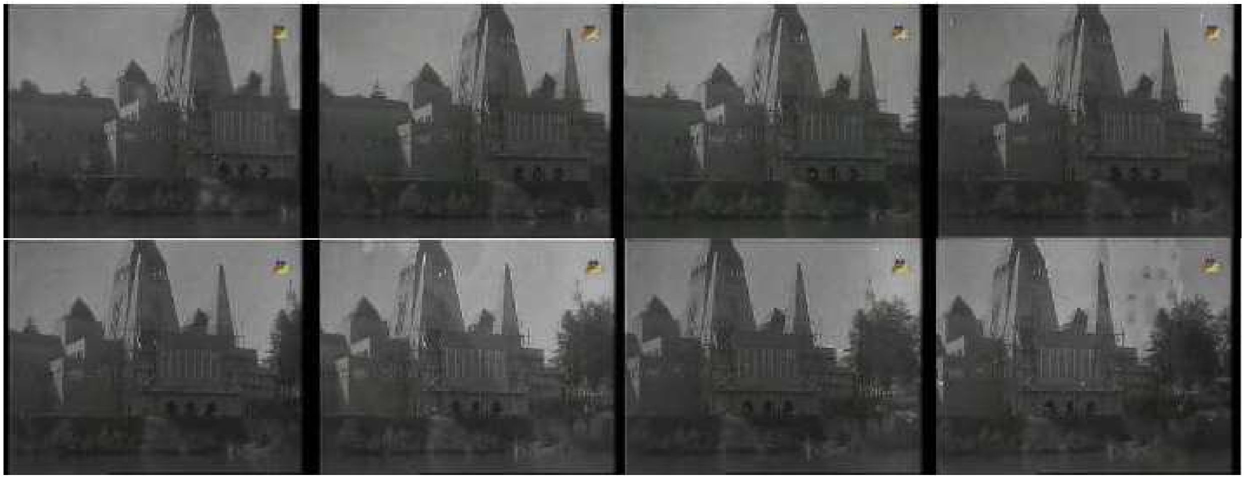


Figure 109. Frame from the film footage: “Nei cantieri dell’esposizione”, Museo del Cinema di Torino and I-media-cities, 1911.

The sources that have preserved traces of the International Exposition of 1928 are the Dezzutti and Melis collections of the historical archive of architecture library of Politecnico di Torino, and the magazine "Domus" (No. 9, September 1928) from which it was possible to recover descriptive texts of the pavilions, measurements, surveys, photos and technical drawings of the projects.

In addition, the Cinema Museum of Turin, in collaboration with the I-Media-Cities platform, preserves the film of the movie "Torino 1928" (Luis Bogino, end 1920s - early 1930s), in which the pavilions appear. The film was shot with a Cine Kodak videocamera and has the following characteristics in Table 18.

Table 18. Technical feature of the film footage "Torino 1928".

Colour	Black & White
Gauge	16 mm
Focal length	25 mm
Digital format Resolution	444x360 pixels

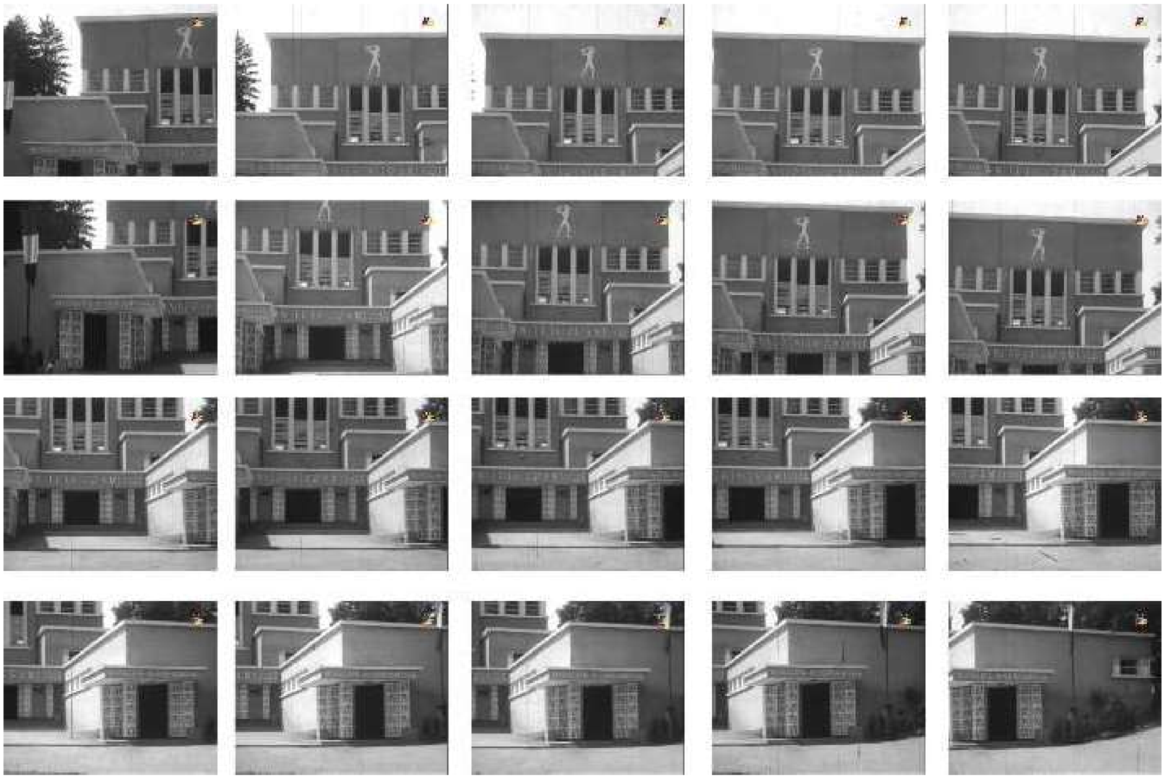


Figure 110. A selection of the frames extracted from the video “Torino 1928” in which the pavilion of “Mines and Ceramics” appeared, Museo del Cinema di Torino and I-media-cities, 1928.

Frames from the footage were extracted and processed implementing the photogrammetric pipeline (Figure 111). The results are shown in Figure 112 and 113.

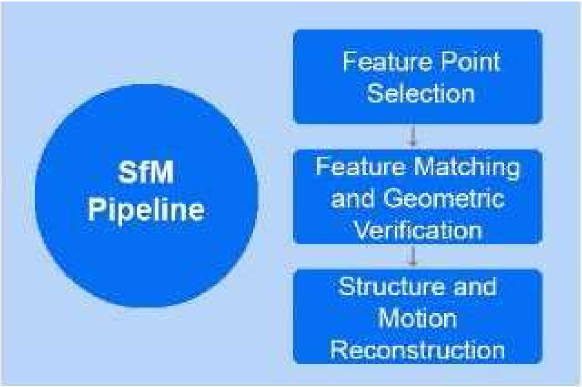

SfM Pipeline
Feature Point Selection
Feature Matching and Geometric Verification
Structure and Motion Reconstruction

Figure 111. A focus on the third step of the general workflow proposed in this dissertation in which the three steps of the photogrammetric pipeline are reported.

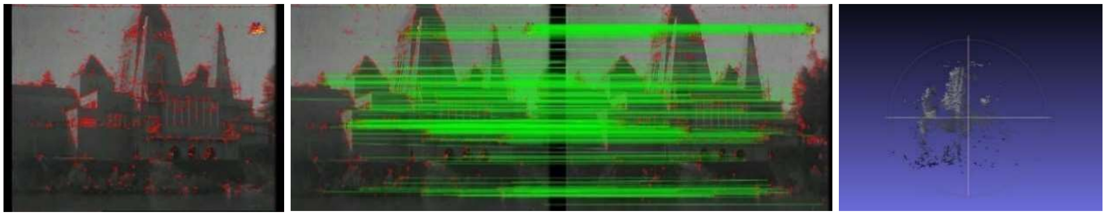


Figure 112. Feature detection and extraction, Feature matching and geometric verification, Point cloud resulted from structure and motion reconstruction of the Hungarian pavilion from the film footage “Nei cantieri dell’esposizione”.

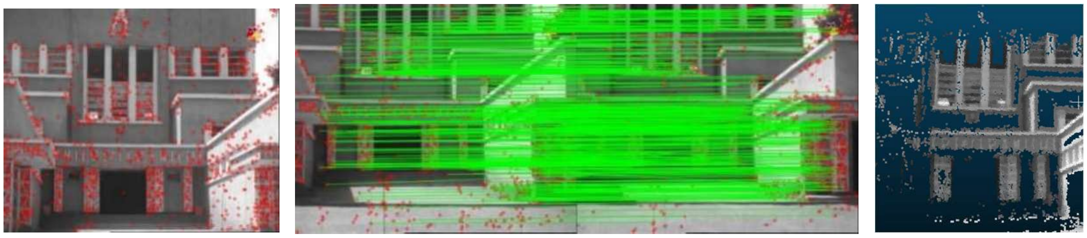


Figure 113. Feature detection and extraction, Feature matching and geometric verification, Point cloud resulted from structure and motion reconstruction of “Mines and Ceramics” pavilion from the film footage “Torino 1928”.

For the precision analysis, the same methodology previously explained was used, and the values of the Final Cost from the bundle adjustment report of the process were exanimated and compared with a benchmark of the maximum metric quality reachable by implementing photogrammetry on videos.

For the Hungarian pavilion the benchmark’s values were chosen according to the Left/Right Motion or Trucking camera motion and a taking distance of 85 m. Consequently the GSD for trucking benchmark is 7.97 [cm/px], and the GSD calculated for the pavilion is 16.52 [cm/px], both calculated on the point closest to the camera.

While for the “Mines and Ceramics” pavilion the benchmark’s values were chosen according to the Left/Right Motion or Trucking camera motion and a taking distance of 25 m. Consequently the GSD for trucking benchmark is 2.34 [cm/px], and the GSD calculated for the pavilion is 3.60 [cm/px], both calculated on the point closest to the camera.

All values of Final Cost were used for the estimation of the Mean and the Standard Deviation and reported in the following graphs to analyse the trend of the data and the comparison with the benchmark.

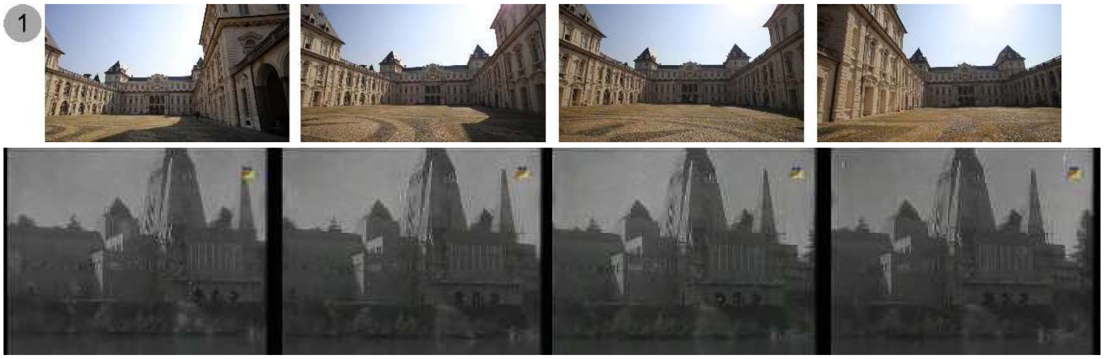


Figure 114. Frames extracted from the video sequences 1 of the benchmark of the façade of the courtyard of the Valentino Castle shot with trucking1 camera motion at a distance of 38 m, compared with the frame extracted from the video “Nei cantieri dell’esposizione”.

Table 19. Values of the benchmark, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding trucking camera motion case. The Trucking1 case, chosen as reference in this processing concerning the Hungarian pavilion for the comparison with the results of the frame extracted from the video “Nei cantieri dell’esposizione”, is highlighted in red.

Camera motion	Mean	Standard Deviation	Min Residual	Max Residual	Distance	GSD	Mean	Standard Deviation	Min Residual	Max Residual
	[px]	[px]	[px]	[px]	[m]	[cm/px]	[cm]	[cm]	[cm]	[cm]
Trucking1	0.6	0.1	0.3	0.9	85.0	8.0	5.0	0.9	2.3	7.5
Trucking2	0.7	0.1	0.2	0.8	65.0	6.1	3.9	0.5	1.4	4.7
Trucking3	0.7	0.1	0.3	0.8	45.0	4.2	2.9	0.4	1.2	3.4
Trucking4	0.7	0.1	0.2	1.1	25.0	2.3	1.6	0.3	0.6	2.6

Table 20. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals of the results of the photogrammetric processing of the frame extracted from the video “Nei cantieri dell’esposizione”, compared with the Trucking1 case of the benchmark.

Case	Mean	Standard Deviation	Min Residual	Max Residual	Distance	GSD	Mean	Standard Deviation	Min Residual	Max Residual
	[px]	[px]	[px]	[px]	[m]	[cm/px]	[cm]	[cm]	[cm]	[cm]
Benchmark	0.6	0.1	0.3	0.9	85.0	8.0	5.0	0.9	2.3	7.5
Case study	0.4	0.1	0.2	0.5	85.0	16.5	5.8	1.2	3.1	8.3

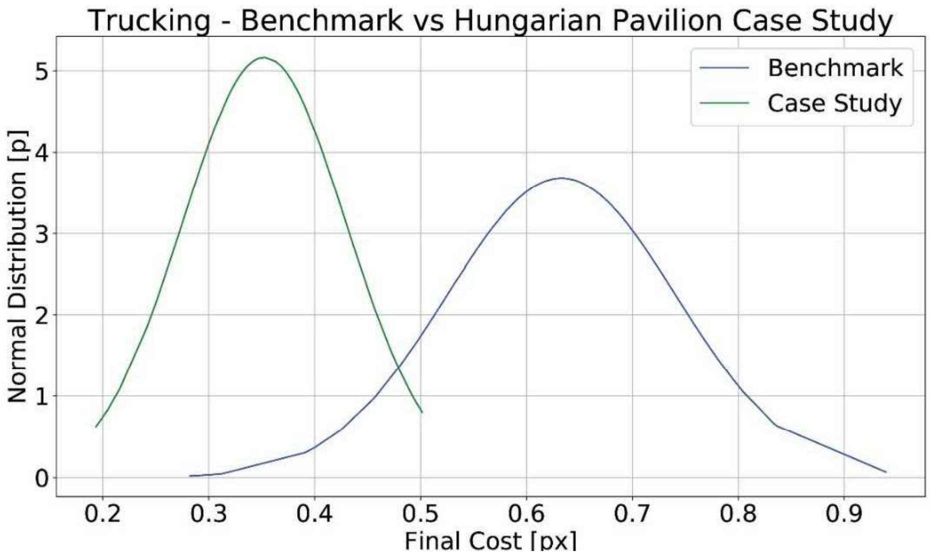


Figure 115. Comparison of Normal Distribution of the Residual value between benchmark tilting2 case and case study of the Hungarian pavilion.

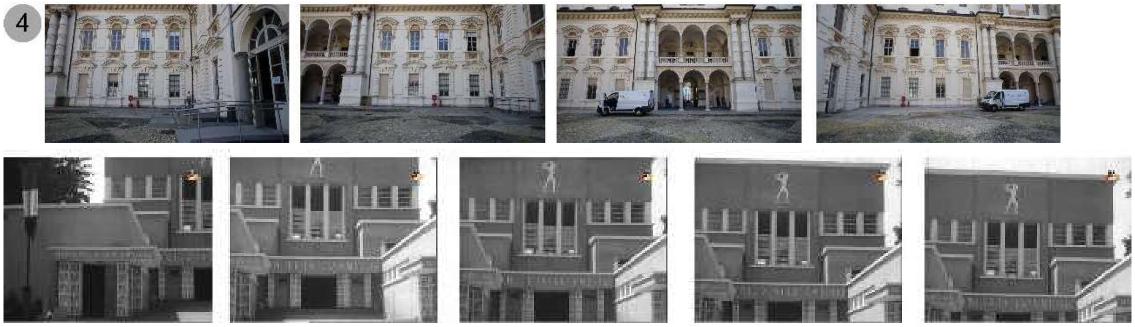


Figure 116. Frames extracted from the video sequences 4 of the benchmark of the façade of the courtyard of the Valentino Castle shot with trucking4 camera motion at a distance of 38 m, compared with the frame extracted from the video “Torino 1928”.

Table 21. Values of the benchmark, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals for each photogrammetric processing, according to the corresponding trucking camera motion case. The Trucking4 case, chosen as reference in this processing concerning the “Mines and Ceramics” pavilion for the comparison with the results of the frame extracted from the video “Torino 1928”, is highlighted in red.

Camera motion	**Mean**	**Standard Deviation**	**Min Residual**	**Max Residual**	**Distance**	**GSD**	**Mean**	**Standard Deviation**	**Min Residual**	**Max Residual**
	[px]	**[px]**	**[px]**	**[px]**	**[m]**	**[cm/px]**	**[cm]**	**[cm]**	**[cm]**	**[cm]**
Trucking1	0.6	0.1	0.3	0.9	85.0	8.0	5.0	0.9	2.3	7.5
Trucking2	0.7	0.1	0.2	0.8	65.0	6.1	3.9	0.5	1.4	4.7
Trucking3	0.7	0.1	0.3	0.8	45.0	4.2	2.9	0.4	1.2	3.4
Trucking4	0.7	0.1	0.2	1.1	25.0	2.3	1.6	0.3	0.6	2.6

Table 22. Values, expressed in pixel and in centimetre, of the Mean, Standard Deviation, Minimum and Maximum of the Residuals of the results of the photogrammetric processing of the frame extracted from the video "Torino 1928", compared with the Trucking4 case of the benchmark.

Case	**Mean**	**Standard Deviation**	**Min Residual**	**Max Residual**	**Distance**	**GSD**	**Mean**	**Standard Deviation**	**Min Residual**	**Max Residual**
	[px]	**[px]**	**[px]**	**[px]**	**[m]**	**[cm/px]**	**[cm]**	**[cm]**	**[cm]**	**[cm]**
Benchmark	0.7	0.1	0.2	1.1	25.0	2.3	1.6	0.3	0.6	2.6
Case study	0.4	0.1	0,02	2.9	25.0	3.6	1.5	0.5	0.1	10.5

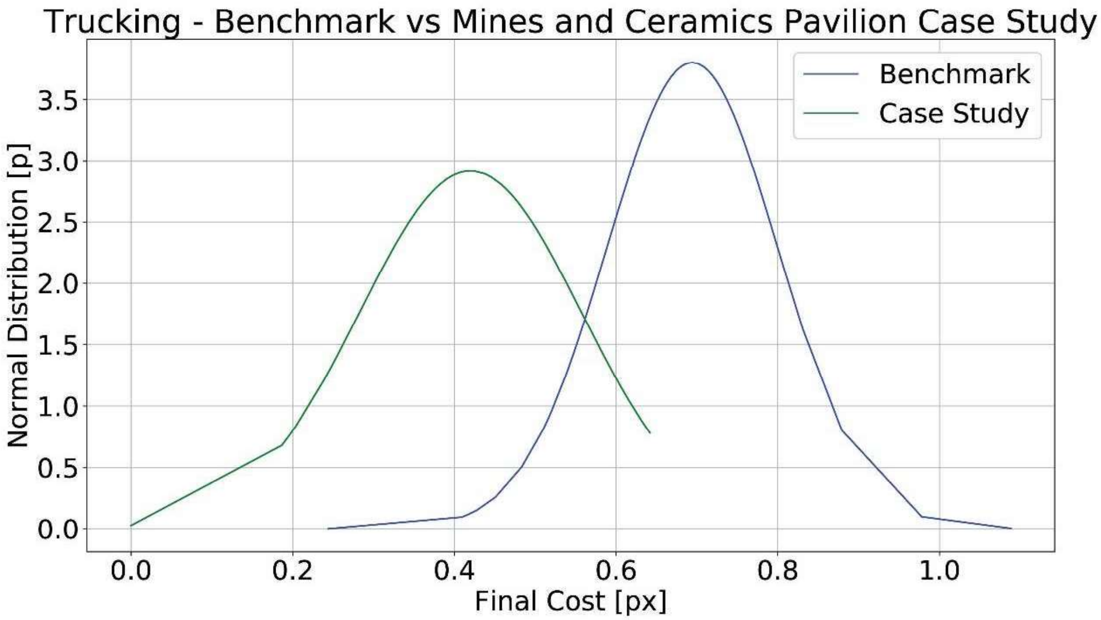


Figure 117. Comparison of Normal Distribution of the Residual value between benchmark tilting2 case and case study of the "Mines and Ceramics" pavilion.

To scale the model od the "Mines and Ceramics" pavilion only some measures from design drawings were available. They were chosen and inserted in the point clouds, considering the limitations due to the fact that probably after the constructions some variations occurred. However, the advantage is that the projects give fundamental and reliable information about the dimensions of the building, useful for the creation of a 3D model and following studies.

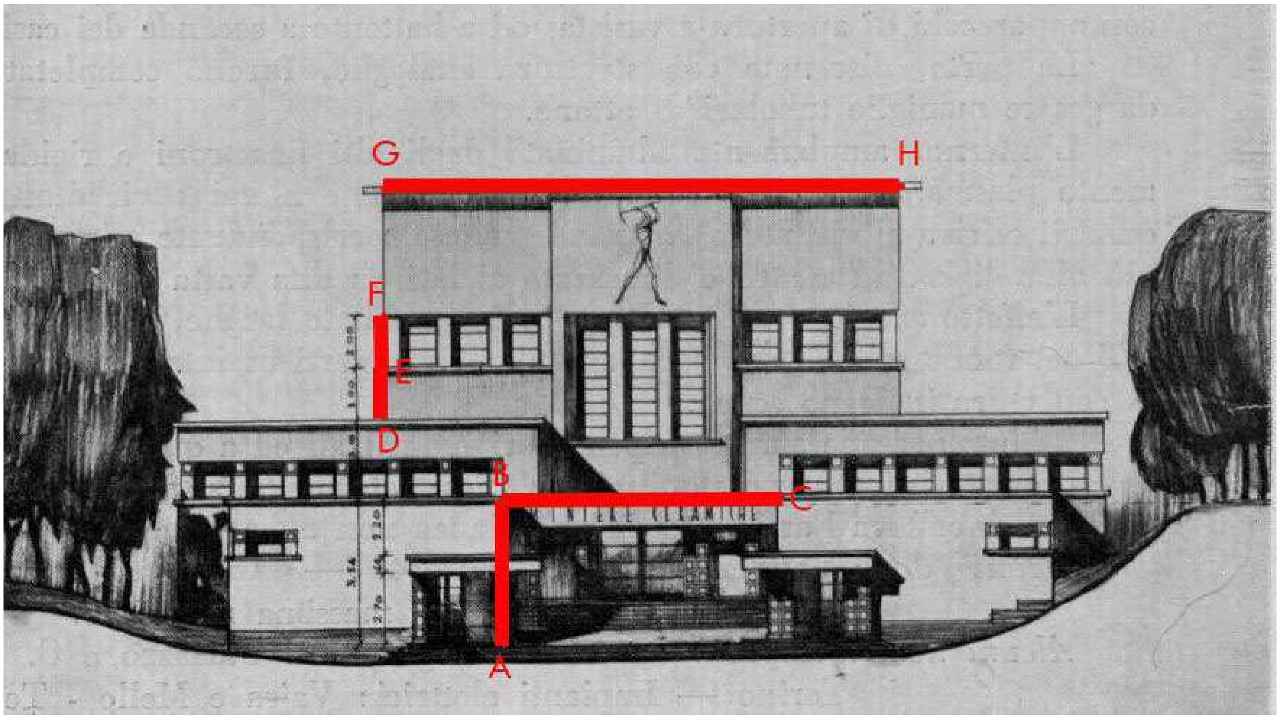


Figure 118. Measure extracted from the project drawing of the "Mines and Ceramics" pavilion. The distance AB was used to scale the model because present in the metric description of the pavilion.

Table 23. Residuals values of distance measured on the project drawing and the point cloud of the "Mines and Ceramics" pavilion obtained from the photogrammetric process of video frame.

Distance	**Project drawing**	**Point Cloud**	**Residuals**
	[m]	**[m]**	**[m]**
AB	5.54	5.47	0.07
DE	1.00	0.98	0.02
EF	2.00	1.88	0.12
GH	20.00	19.36	0.64
BC	10.60	9.75	0.85

The obtained differences show a good quality of the survey also by considering that the reference distances are the one extracted from a design drawing, and it is acceptable that some of them were not correctly realized on the site. Finally, the results, within tolerances described just now, could be considered adequate.

References

Baltard, V., 1863. Monographie des Halles centrale de Paris. Archives de Paris, ATLAS 97: INHA.

Condorelli, F. and Rinaudo, F.: Cultural Heritage reconstruction from historical photographs and videos, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2, 259-265, https://doi.org/10.5194/isprs-archives-XLII-2-259-2018.

Condorelli, F., Rinaudo, F., Salvadore, F., Tagliaventi, S., 2020. A match-moving method combining AI and SfM algorithms in historical film footage, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B2-2020, 813–820, https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-813-2020, 2020.

Condorelli, F., Rinaudo, F., Salvadore, F., Tagliaventi, S., 2020. A Neural Networks Approach to Detecting Lost Heritage in Historical Video, ISPRS Int. J. Geo-Inf. 2020, 9(5), 297; https://doi.org/10.3390/ijgi9050297.

Condorelli, F. and Rinaudo, F., 2019. Processing historical film footage with Photogrammetry and Machine Learning for Cultural Heritage documentation. In MM '19: 2019 ACM Multimedia Conference Proceedings, October 2019, Nice, France. ACM, NY, USA. 8 pages. https://doi.org/10.1145/3347317.3357248.

Condorelli, F., Rinaudo, F., Salvadore, F., and Tagliaventi, S., 2019. Architectural Heritage recognition in historical film footage using Neural Networks, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W15, 343–350, https://doi.org/10.5194/isprs-archives-XLII-2-W15-343-2019, 2019.

Cornaglia P., 2001. A magyar pavilon az 1911 - es Torinói Világkiállításon - Il padiglione ungherese all'Esposizione Internazionale di Torino del 1911, Pavilon, OMvH Magyar Èpítészeti Múzeum Pavilion Alapítvány, Budapest, Hungary, 79-96.

Della Coletta, C., 2006. World's Fairs Italian-Style: The Great Exhibitions in Turin and Their Narratives, 1860-1915. University of Toronto Press, Toronto, ON, Canada.

De Moncan, P., 2012. Baltard, Les Halles de Paris 1853-1973, Les Editions du Mécène, ISBN-10 : 2907970992.

Hu, B., Lu, Z., Li, H., Chen, Q., 2014. Convolutional neural network architectures for matching natural language sentences. Adv. Neural Inf. Process. Syst. Nips 2014, 27, 1–9.

Kalal, Z., Matas, J., Mikolajczyk, K, 2010. P-n learning: Bootstrapping binary classifiers by structural constraints. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1–8.

Lemoine, B., 1980. Les Halles de Paris, la storia di un luogo, gli alti e bassi di una ricostruzione, il susseguirsi di progetti, l'architettura di un monumento, la sfida di una "Città", Éditions l'Équerre.

Medecine de France N°226, Paris: Olivier Perin Éditeur MCMLXXI 1971.

Meurgey, J., 1926. Histoire de la paroisse Saint-Jacques de-la-Boucherie; Bibliothèque de l'École des chartes: Paris, p. 347.

O'Connel, L. M., 2001. Afterlives of the Tour Saint Jacques: Plotting the perceptual history of an urban fragment, Journal of the Society of Architectural Historians (2001) 60 (4): 450–473, https://doi.org/10.2307/991730.

Pinon, P., 2012. Paris pour mémoire - Le livre noir des destructions haussmanniennes, Relié.

Tamborrino, R., 2005. Parigi nell'ottocento, Marsilio, Venezia, p. 69.

Thomine-Berrada, A., 2012. Victor Baltard : Architecte de Paris, Gallimard, ISBN-10: 2070447367.

Chapter 5

Conclusions

5.1 Restating the aims of the dissertation

This dissertation has demonstrated the potentialities of historical images for the metric documentation of Cultural Heritage in the photogrammetric field combined with Artificial Intelligence. Photogrammetry and Deep Learning were used in an innovative way to extract metric information of historic buildings from historical photographs and film footage for their 3D virtual reconstruction.

The study underlines the potentialities of the use of historical images in the field of historical studies. Libraries, archives and repositories today provide a large amount of digital data, and as a consequence, problems such as differences in data quality and organization arise.

By using image collections, an innovative method for analyzing historical images and expanding the use of photo and video archives is proposed. Through the analysis of image search strategies for the virtual reconstruction of Cultural Heritage, a systematic classification of images suitable for this purpose has been carried out.

Furthermore, a new workflow was proposed to automatically select images suitable for photogrammetric processing of historical images in order to document Cultural Heritage with certified metric quality.

With the implementation of open-source Structure-from-Motion algorithms, it is possible to obtain the 3D virtual reconstruction of the monuments and their metric information.

In the first phase of the workflow, in order to make automatic the search for a specific monument to be documented, an algorithm has been developed for the detection of architectural heritage in historical images through Deep Learning. This algorithm made it possible to identify the frames in which the monument appeared without having to manually examine the various videos in the archive, thus reducing time and increasing efficiency, and to process them by photogrammetry.

The use of object detection has proven to be a good solution for the automatic recognition of architectural heritage in historical videos, especially because it allows the extraction of the coordinates of the bounding boxes that locate the monument within the frame. The experiment focused on reducing the human effort to detect the wanted object and increasing the efficiency of the operator's work in the archive. To achieve this goal, Deep Learning algorithms were identified as potential solutions to reduce the time needed to search for monuments in the video records of historical archives.

The second step was the automatic identification of camera movement in order to select images suitable for photogrammetry. Following the evolution of bounding boxes detected by the Neural Networks, the algorithm identifies camera movements in significant categories. In particular, this strategy is used to detect tilting and trucking camera motions, which are suitable for photogrammetry and are very common in historical videos. Thanks to its structure, the algorithm can work even when the track is very short. The video frames are first grouped in frame clusters according to image similarity criteria. Then, for each frame cluster, the camera motion category is evaluated. This is particularly useful in photogrammetry, where each group of frames can be used separately to increase the positive completion of the procedure. In fact, photogrammetric reconstruction is bound to fail when images from different scenes or videos are used in the same process. The proposed algorithm includes some parameters that, if properly selected, can reach an overall accuracy of up to 80% in identifying camera movements. This accuracy is obtained by minimizing the misinterpretation of camera oscillations, due to the poor quality of the camera used to take historical film footage and the absence of a tripod, and by correctly setting the image similarity criterion used to group images.

The third step is to process the image extracted from the film using open-source SfM algorithms. A method is experimented with to manually extract feature points in photogrammetric processing of historical images in order to use them in the matching phase of the reconstruction and to ensure their presence in the resulting point cloud, even if poor. This allowed the metric evaluation of the quality of the results during the last phase by comparing point clouds of different density and resolution, which certifies the quality of the 3D reconstruction.

5.2 Originality and findings of the research

The originality of the proposed workflow lies in the improvement of the photogrammetric pipeline by using Deep Learning algorithms. In fact, the recognition of the monument in the video was inserted as the first step of the photogrammetric reconstruction.

The research also showed how Neural Networks can be effectively trained for historical monument search. In particular, firstly the algorithm was tested on two different architectures. The first was the Tour Saint Jacques for the tuning of the networks in the best situation of a heritage that still exists but has undergone many changes. Therefore, a large amount of material was available to test the algorithm and obtain a metric comparison and test the potential of the approach. The second case study was Les Halles selected to test the algorithms on a real case of a destroyed architecture. The performance of the network was evaluated using different datasets, depending on the different conditions found in the historical data. According to the appropriate metrics of the cases in question, the quality of the results is encouraging, both in terms of the human time saved and the results obtained.

The experiments were conducted using High-Performance Computing (HPC) clusters IBM Power9 with NVIDIA v100 by the research centre CINECA. Thanks to the use of this hardware, the results showed that the reduction of the time required to process an image during the training stage is about 95% (0.3 s/image VS 9 s/image of a mid-range laptop).

The metric quality of the 3D models obtained from the photogrammetric processing of historical images was also evaluated against a new benchmark creating with the purpose of establishing the maximum metric quality reached with these kinds of reconstruction.

The results of this study are very encouraging. Indeed, working with historical images has inherent difficulties such as the lack of important information about the camera, the quality of the film used to shoot them, and the lack of an accurate metric reference when the monument is lost. The results presented in this thesis show that the presented automatic workflow can be effective even under these critical conditions.

The method has been validated on other case studies chosen in order to verify it, and in most cases the 3D model obtained from historical images provided results of acceptable accuracy, even in cases where the object reconstructed from historical photographs no longer exists.

This evaluation is interesting because shows how to harness historical material available in case of Cultural Heritage destroyed. This compared analysis gives an idea of the reliability of the 3D reconstruction from historical images, considering all the limitations intrinsically present in the primary data.

5.3 Applications and further research

The experimental work presented in this thesis is an investigation into the level of quality of results it is possible to reach when processing historical images from archives.

There are several areas where this study makes an original contribution in the field of Cultural Heritage. Although the analysis of the case studies was treated from a technical point of view, the findings of this thesis have significant implications for the understanding of historical considerations.

Besides the creation of a new tool for searching for historical material in archives thanks to the automation of a manual task and the improvement of the photogrammetric process by selecting the right material for the application, the research has a great impact on the protection and valorisation of Cultural Heritage from a different point of views.

The virtual reconstruction of transformed or lost Cultural Heritage allows historians and architects to explore how it was in the past and to understand its development and the original state of buildings and urban environments.

In order to enhance the archive and innovatively exploit archival resources, the use of Deep Learning actually strengthened known methods of documenting lost heritage. It gave an impact not only on improving the existing images archive platform creating a more efficient and accurate system for the users of digital resources (scholars, educators, student, museum etc.) but also offered some important insights into the management and organization of historical information and the protection of the past.

This information is extremely useful to make decisions and interventions on the heritage, for management, restorations works and structural analysis.

Despite its limitations, the study certainly contributes to greater awareness in the valorisation of Cultural Heritage data and should be repeated using different datasets and imaging conditions.

The approach described in this work can be applied to different historical monuments. Further research is needed to evaluate the effectiveness of the experimental methodology and to extend its application to other case studies, especially lost heritage. In particular, it would be interesting to apply the procedure to other destroyed monuments for which 3D reconstruction from historical videos is the only possible option.

Another interesting future extension of this study could tackle the complexity of historical data. Further research will expand the discussion on open issues in historical archives and provide references for possible solutions. The development of a standard structure for metadata concerning historical images, for example, will allow the classification and the link of collections across different database and institutions.

From a technical point of view, there are many aspects which can be further investigated. In particular, future studies should analyse a complete classification of existing Neural Networks according to the different applications. Furthermore, the introduction of other types of Neural Network models is another possible solution that should be tackled to investigate how the accuracy of recognition can be modified. The experimentation of an open-source cloud environment for the training of the networks (such as Kubernetes - Kubernetes.io -, last access

February 2021) could be implemented in order to use hardware different from the HPC.

Another important development could be the refinement and improvement of the algorithm for camera movement, e.g. by developing an automatic determination of camera movements in the time intervals in which the object was detected. In principle, this would allow automation of the entire photogrammetric pipeline and could therefore be a particularly interesting field of research. In addition, the same methodology could be applied to newer videos and allow access and recording of information of daily life. For this reason, further research could evaluate the effectiveness of the experimental methodology, here tested on the heritage field, in other areas, such as UAVs, structural analysis and computer vision.

From the end user's point of view, another area of development could be the creation of digital services in order to simplify the use of Deep Learning for a potential non-expert user. The development of an intuitive interface that allows the automation of the most complex steps of the process would be a great help to improve the usability of the workflow.

Another useful digital service could be the use of the outputs of the 3D reconstruction for the development of VR and AR apps to improve the involvement and interest of tourists in the collections in a more engaging way, also remotely due to the COVID-19 situation.

5.4 Importance of this research for the present

As the work of an archaeologist that recovers and analyses the material traces left from the culture of the past, this research is a first attempt to bring to light traces of the memory through the reconstruction of fragments of Cultural Heritage that appear in historical images. Linking different fragments of photographs or footage in which a place appears in a specific time interval or different historic period, could help to reconstruct the memory of heritage throughout history.

Taking advantages of photogrammetry and Artificial Intelligence technologies allowed the identification and the virtual reconstruction of remaining traces of heritage monuments and parts of a city that have been lost or changed over time. However, the potentialities of the method go beyond the simple process of documenting something real that existed in the past.

The strength of the method lies in creating the information and knowledge base for the future generation. In fact, the pictures and the videos that are taken every day simply by walking through a city can be used in the future to reconstruct the Cultural Heritage.

Finding new ways to re-discovering the past and dealing with the historical material that will become a memory for the future is the main challenge faced in this research.