Affine scale space for viewpoint invariant keypoint detection

Original
Affine scale space for viewpoint invariant keypoint detection / Zhao, B., Lepsoy, S., Magli, E.. - ELETTRONICO. - (2015), pp. 1-6. (2015 IEEE International Workshop on Multimedia Signal Processing 2015) [10.1109/MMSP.2015.7340860].

Availability:
This version is available at: 11583/2638703 since: 2016-04-01T12:47:43Z

Publisher:
IEEE

Published
DOI:10.1109/MMSP.2015.7340860

Affine scale space for viewpoint invariant keypoint detection

Biao Zhao[1], Skjalg Lepsoy[2], Enrico Magli[3]

[1,3]*Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino*
Corso Duca degli Abruzzi, 24 - 10129 Torino, ITALY
biao.zhao@polito.it[1]
enrico.magli@polito.it[3]

[2] *Joint Open Lab, Telecom Italia*
Corso Duca degli Abruzzi,24 -10129 Torino, ITALY
skjalg.lepsoy@telecomitalia.it[2]

***Abstract*—The research of affine scale space is to create a more general approach to the affine invariant image scale representation by modifying the corresponding Gaussian filters in order to cope with the specific change of view point. It has the purpose to retain a linear relationship with the transiting of the view point. With this linear relationship, the affine scale space could be established as a more general approach for the affine invariant image retrieval, including affine feature detection and affine feature descriptor. The scope of this paper is to discuss the accessible to the affine scale space, its performance and a practical implementation to construct it in order to cope with the high complexity brought in by the scale space and the affine adaptation.**

I. Introduction

The proliferation of digital cameras in the smartphone and consumer-level products is producing an explosion of media data, including videos and images. In front of these massive amounts of visual data, it is necessary to develop some efficient visual search techniques for multimedia browsing, searching and retrieval. Most traditional and common methods of visual search utilize metadata such as captioning, keywords, or descriptions of the media so that retrieval can be performed over the annotation words. Manual annotation is time-consuming, laborious and expensive; to relieve the conflict between low efficient media annotation and rapid increasing of media data, a large effort has been made to automatically annotate the visual data by analysing its content, leading to the content-based visual search techniques [1].

Content-based means that the retrieval is based on the contents rather than the metadata such as keywords, or tags associated with the images or videos. Content of the visual data can refer to colours, shapes, textures, edges, or any other inherent feature of the image. A valid and robust visual search system should based on features that are stable and robust enough to the fluctuation of the scene, including the effect of illumination, exposure, partial occlusion and scale.

The most commonly applied scale invariant feature detectors are generally based on Laplacian of Gaussian (LoG) [2], which detects the local extrema on LoG filtered images. Usually, LoG results in strong positive responses for dark blobs and strong negative responses for bright blobs. In order to automatically capture blobs in the image domain, where the specific scale is not known in advance, a multi-scale approach is therefore necessary, which inspires the creation of scale space.

A scale space can be defined as a collection of several pre-smoothed images by different sized Gaussian kernels. In general, scale space theory is a framework for multi-scale signal representation for handling image structures at different scales, by employing a one-parameter family of smoothed images. This framework provides a scale-invariant representation, which is necessary for dealing with the size variations that may occur in image data. Indeed, real-world objects, in contrast to idealized mathematical entities, may appear in different ways depending on the scale of observation [3].

A highly useful property of scale space is that feature detection can be made scale invariant, by performing automatic scale selection based on normalized derivatives. Based on this multi-scale image representation, several visual search techniques have been proposed, including Scale Invariant Feature Transform (SIFT) [4] and A Low-dimensional Polynomial detector (ALP) [5].

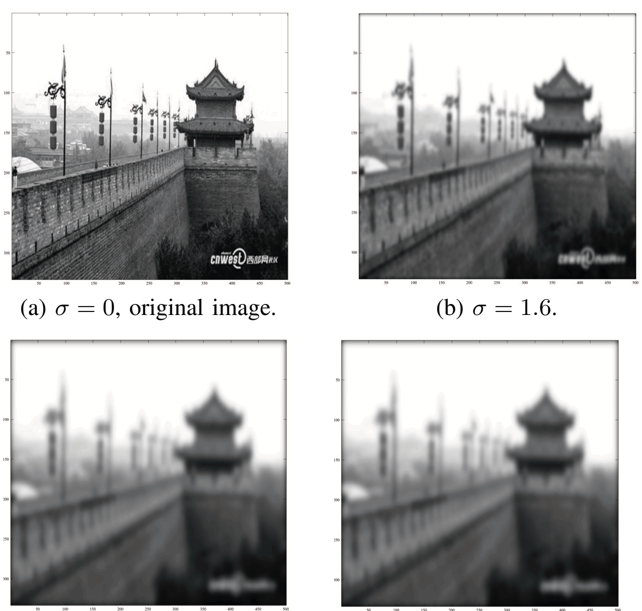
On the other hand, robustness to different view points is also an important criterion to evaluate the performance of a visual search technique. Compared to some other visual search techniques, SIFT is partially robust to affine transformation, but not enough to be able to match images taken from very different view-points.

The performance evaluation by Mikolajczyk and Schmid [6] presents a comparison of different visual search techniques in terms of their robustness to view point changes. It showed that SIFT has just a little above 50% correct matching ratio under large viewpoint changes, which is rather low referring to typical image matching precision.The evaluation reveals a very important issue, which has not been properly addressed in the existing literature, namely the robustness of visual search algorithms to view-point changes. In the real world, a scene may be acquired from several different view points, resulting in a large number of images with quite different perspectives for the same content. Without invariance or robustness to affine transformations, the application of content-based visual search will be seriously limited. In addition, some practical applications such as 3D object reconstruction, object recognition and architecture retrieval rely on accurate visual search of multiplexed planar objects. Without an accurate and fully viewpoint invariant image retrieval system, such application can not be carried out.

In order to address this problem, in this paper we propose an affine scale space as a more general approach to the scale-invariant image representation in order to compensate for the non-linear relationship brought in by the change of view point. The affine scale space can be generated by steering the Gaussian filters to the specific affine transformations if the change of view point is previous known. In this paper, we also propose a practical structure for real time implementation. The affine scale space is a forward model, which can be used to predict what will happen to an image under a different view point. The performance of this affine scale space will be evaluated by calibrating its feature detection precision based on the affine scale space in an affine invariant way.

II. Scale Space and its Implementation

Scale space is the fundamental theory of multi-scale signal representations, handling image structures at different scales by representing an image as a one-parameter family of smoothed images. The scale-space representation, parametrized by the size of the smoothing kernel, is used for suppressing fine scale structures [7].



(a) $\sigma = 0$, original image. (b) $\sigma = 1.6$.

(c) $\sigma = 1.6 \times 2$. (d) $\sigma = 1.6 \times 2\sqrt{2}$.

Fig. 1. A typical scale space, including the images smoothed with different size of Gaussian filter.

For a given image $I(x, y)$, its scale-space representation is given by a family of images $L(x, y, \sigma)$, smoothed by two dimensional Gaussian kernels, whose parameter is defined according to the kernel size:

$$g(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{1}$$

such that

$$L(x, y; \sigma) = g(x, y; \sigma) * I(x, y) \tag{2}$$

The scale parameter σ is the standard deviation of the Gaussian filter. Figure 1 represents a typical scale space. Increasing σ, the corresponding image in the scale space is smoothed with a larger Gaussian kernel and hence it contains fewer details.

The scale space representation contains interesting image structures at all the scales. In order to capture the structure at the corresponding scale, an appropriate metric should be found. The local extrema in the scale space are a good choice for this metric. Thus, it can apply the Laplacian operator to detect the local maximum or minimum by means of normalized derivative on each smoothed image in the scale space. With this Laplacian operator, the automatic feature detection can easily be implemented in the framework of scale space.

The Laplacian operator is defined as,

$$\nabla^2 L = L_{xx} + L_{yy} \tag{3}$$

A practical implementation of scale space relies on the pyramid structure, which allows to obtain a computationally efficient approximation to scale space. There are mainly two types of structure: Gaussian pyramid and Laplacian pyramid.

In the pyramid, the images blurred by different Gaussian kernels will be divided into small groups termed as octave. Between each octave, the image will be sub-sampled as the input image for the next octave in order to reduce the exponentially increased Gaussian kernel. In this way, the processing can be greatly accelerated.

The common procedures to construct the Gaussian pyramid is as follows: the original image will be convolved with a low-pass Gaussian filter as the first blurred image in the scale space. This blurred image is repeatedly convolved with different size of Gaussian filters until all the blurred images in the first octave have been created. Then, one of these images will be sub-sampled as the initial image for the next octave. The images of the next octave will also be created by the same series of low-passed Gaussian filters. This smooth-sub-sampling will be recursively operated to create the whole scale space.

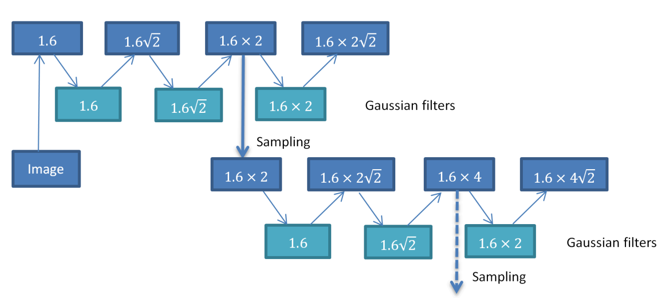


Fig. 2. ALP pyramid structure.

Figure 2 presents a typical pyramid structure which has been used in ALP. We can notice that only three Gaussian filters have been recursively applied to construct the whole Gaussian pyramid.

III. Affine scale space

Most works on multi-scale representations focus on the definition of isotropic scale space, characterized by equal behaviour in all directions [7]. But that kind of scale space is not compatible with non-isotropic image structures generated by a perspective affine transformation, which models well an image taken from a different view point. Our proposed solution to this issue is a new scale space that accounts for linear geometric transformations. This new scale space is termed ”affine scale space”.

Assume that two images I_1 and I_2 have been taken from different view point, which can be modelled as an affine transformation:

$$\begin{gathered} I_1(\xi) = I_2(\eta), \text{ where } \eta = A\xi. \\ \text{Thus, } I_1(\xi) = I_2(A\xi). \end{gathered} \tag{4}$$

Both ξ and η are two dimensional vectors and A is a 2×2 matrix. The traditional Gaussian scale space can be defined as

$$L(x, y, \sigma) = g(x, y, \sigma) * I(x, y). \tag{5}$$

It can also be expressed in the form of vector by defining $\xi = (x, y)$,

$$L(\xi, \sigma) = g(\xi, \sigma) * I(\xi). \tag{6}$$

Thus,

$$\begin{aligned} L_1(\xi;\sigma) =& g(\xi;\sigma) * I_1(\xi) = g(\xi;\sigma) * I_2(A\xi) \\ L_2(\eta;\sigma) =& g(\eta;\sigma) * I_2(\eta) \end{aligned} \tag{7}$$

It is clear that

$$\begin{aligned} g(\xi;\sigma) * I_2(A\xi) \neq& g(\xi;\sigma) * I_2(\xi) \\ L_1(\xi;\sigma) \neq& L_2(\xi;\sigma) \end{aligned} \tag{8}$$

Equivalently,

$$\begin{aligned} g(\xi;\sigma) * I_2(A\xi) \neq& g(A\xi;\sigma) * I_2(A\xi) \\ g(\xi;\sigma) * I_1(\xi) \neq& g(\eta;\sigma) * I_2(\eta) \\ L_1(\xi;\sigma) \neq& L_2(\eta;\sigma) \end{aligned} \tag{9}$$

This equation shows that geometric linear relationship will no longer exist between the classical scale spaces of the images taken under different view points.

A reasonable approach to handle this deformed structure is to generate the corresponding scale space also by a deformed Gaussian kernel. A classical Gaussian kernel can be expressed as

$$g(\xi;\sigma) = \tfrac{1}{2\pi\sigma^2} e^{-\frac{\xi^T\xi}{2\sigma^2}}. \tag{10}$$

Since $\eta = A\xi$, then we have

$$\begin{aligned} \xi =& A^{-1}\eta \\ g(\xi;\sigma) =& \tfrac{1}{2\pi\sigma^2} e^{-\frac{(A^{-1}\eta)^T(A^{-1}\eta)}{2\sigma^2}} \\ g(\eta;\sigma)_{af} =& \tfrac{1}{2\pi\sigma^2} e^{-\frac{\eta^T(AA^T)^{-1}\eta}{2\sigma^2}} \end{aligned} \tag{11}$$

With the definition of covariance matrices $\Sigma_s = A\sigma^2 A^T$ we obtain

$$g(\eta;\Sigma_s)_{af} = \tfrac{1}{2\pi\sqrt{det\Sigma_s}} e^{-\frac{\eta^T\Sigma_s^{-1}\eta}{2}}. \tag{12}$$

If $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, the linear scale space becomes a special case of affine scale space.

This is the affine Gaussian kernel extended from its classical Gaussian version. Employing this type of filter, the corresponding affine Gaussian scale space can be constructed.

As we have discussed, local feature extraction can be achieved by means of differential derivation on the set of scale space. Under an affine transformation, the derivative of the affine Gaussian scale space can also be obtained by adapting the corresponding Laplacian operation for the affine transformation. In practice, Laplacian operator is usually simplified as a 3×3 Laplacian filter, which is rather difficult to be further affine transformed. Thus it is reasonable to employ the Laplacian of Gaussian (LoG) for the affine adaptation.

The two dimensional Laplacian operator is given by

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \tag{13}$$

The expression of LoG can be derived as:

$$\begin{aligned} L_\Delta(x, y, \sigma) =& \nabla^2 g(x, y, \sigma) \\ =& \frac{\partial^2 g(x, y, \sigma)}{\partial x^2} + \frac{\partial^2 g(x, y, \sigma)}{\partial y^2} \\ =& -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right) e^{-\frac{x^2+y^2}{2\sigma^2}} \end{aligned} \tag{14}$$

Another equivalent expression is given in the vector form,

$$L_\Delta(\xi;\sigma) = -\frac{1}{\pi\sigma^4}\left(1 - \frac{\xi^T\xi}{2\sigma^2}\right) e^{-\frac{\xi^T\xi}{2\sigma^2}}, \tag{15}$$

$$\xi = \begin{pmatrix} x \\ y \end{pmatrix}.$$

The affine LoG filter can also be derived like affine Gaussian filter. Suppose an image transformation is

$$I_1(\xi) = I_2(\eta), \text{ where } \eta = A\xi. \tag{16}$$

Since $\eta = A\xi$, then

$$\xi = A^{-1}\eta \tag{17}$$

$$\begin{aligned} L_\Delta(\eta;\sigma) =& -\frac{1}{\pi\sigma^4}\left(1 - \frac{(A^{-1}\eta)^T(A^{-1}\eta)}{2\sigma^2}\right) e^{-\frac{(A^{-1}\eta)^T(A^{-1}\eta)}{2\sigma^2}} \\ =& -\frac{1}{\pi\sigma^4}\left(1 - \frac{\eta^T(AA^T)^{-1}\eta}{2\sigma^2}\right) e^{-\frac{\eta^T(AA^T)^{-1}\eta}{2\sigma^2}} \end{aligned} \tag{18}$$

and with the definition of covariance matrices $\Sigma_s = A\sigma^2 A^T$ we have

$$L_\Delta(\eta;\sigma) = -\frac{1}{\pi\sigma^4}\left(1 - \frac{\eta^T\Sigma_s^{-1}\eta}{2}\right) e^{-\frac{\eta^T\Sigma_s^{-1}\eta}{2}}. \tag{19}$$

This is how we derive the affine LoG kernel, which can be used to build the affine LoG scale space.

IV. A FEASIBLE STRUCTURE TO BUILD THE AFFINE SCALE SPACE

In practice, constructing a scale space is computationally complex and time consuming. It will be quite appealing if some efficient structure, similar to a pyramid, can be applied to build both the affine Gaussian scale space and affine LoG scale space. In the following, we will propose such a feasible structure.

The fundamental of pyramid structure are two Gaussian axiom properties: semi-group and sub-sampling. The semi-group property of Gaussian axiom can be mathematically expressed as:

$$g(\xi,\sigma_1^2)*g(\xi,\sigma_2^2)=g(\xi,\sigma_1^2+\sigma_2^2) \quad \text{with } \xi=\begin{pmatrix}x\\y\end{pmatrix} \tag{20}$$

Consider an affine transformation which can be mathematically expressed as $\eta = A\xi$. Then,

$$\begin{aligned}\xi &= A^{-1}\eta\\ g(A^{-1}\eta,\sigma_1^2)*g(A^{-1}\eta,\sigma_2^2) &= g(A^{-1}\eta,\sigma_1^2+\sigma_2^2)\end{aligned} \tag{21}$$

So, affine Gaussian scale space also owns the semi-group property.

In the same way, it can be derived that:

$$L_\triangle(A^{-1}\eta,\sigma_1^2)*g(A^{-1}\eta,\sigma_2^2)=L_\triangle(A^{-1}\eta,\sigma_1^2+\sigma_2^2) \tag{22}$$

where $L_\triangle$ stands for a LoG filter.

This is a cascade implementation of LoG simply switching one of the Gaussian filter to LoG filter in the framework of semi-group property, which provides a feasible implementation for the affine adapted LoG scale space. This cascade implementation of LoG has not been properly addressed before. In practice, a LoG scale space can be easily calculated by a Laplacian operation on the corresponding Gaussian scale space. While the Laplacian operation does not provide a feasible way for the affine adaptation, this cascade implementation, based on affine LoG filter, is quite simple and can be used in a straightforward way to construct the affine LoG.

Consider an image in the scale space which can be sub-sampled as

$$\begin{aligned}L(x,y,\sigma^2) &= I(x,y)*g(x,y,\sigma^2)\\ &= I(x,y)*\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}\end{aligned} \tag{23}$$

by sub-sampling,

$$\begin{aligned}L(2x,2y,\sigma^2) &= 4I(2x,2y)*g(2x,2y,\sigma^2)\\ &= 4I(2x,2y)*\frac{1}{2\pi\sigma^2}e^{-\frac{4x^2+4y^2}{2\sigma^2}}\\ &= 4I(2x,2y)*\frac{\frac{1}{4}}{2\pi\frac{\sigma^2}{4}}e^{-\frac{x^2+y^2}{2\frac{\sigma^2}{4}}}\\ &= I(2x,2y)*g(x,y,\left(\frac{\sigma}{2}\right)^2)\end{aligned} \tag{24}$$

or

$$\begin{aligned}L(2x,2y,(2\sigma)^2) &= I(2x,2y)*g(x,y,\sigma^2)\\ L(4x,4y,(4\sigma)^2) &= I(4x,4y)*g(x,y,\sigma^2)\\ &\vdots\end{aligned} \tag{25}$$

In this way, the scale space can be constructed without increasing the size of Gaussian kernels, which would cause increase of computation.

Let $\xi = \begin{pmatrix}x\\y\end{pmatrix}$, then we have,

$$\begin{aligned}L(\xi,\sigma^2) &= I(\xi)*g(\xi,\sigma^2)\\ L(2\xi,(2\sigma)^2) &= I(2\xi)*g(\xi,\sigma^2)\\ \eta = A\xi \quad &\text{so} \quad \xi = A^{-1}\eta\\ L_1(A^{-1}2\eta,(2\sigma)^2) &= I_1(A^{-1}2\eta)*g(A^{-1}\eta,\sigma^2)\\ L_2(2\eta,(2\sigma)^2) &= I_2(2\eta)*g_{af}(\eta,\sigma^2)\end{aligned} \tag{26}$$

where g_{af} is the affine Gaussian filter g under the affine transformation given by A. In this way, the affine scale space can also be sub-sampled to prevent an increasing size of the convolution kernels.

Based on the properties above, we propose our implementation for the affine scale space and affine LoG. Since the properties which the pyramid is based on are also satisfied by affine scale space, theoretically, the same pyramid structure can be completely inherited by the affine scale space. In practice, we can construct the affine scale space by the traditional pyramid structure if we ignore the LoG scale space. In the traditional scale space, the construction of LoG has never been addressed, since the LoG can easily be obtained by the Laplacian matrix applied on the corresponding Gaussian scale space. But in the affine scale space, the corresponding LoG can not be generated by a simple affine Laplacian operation. Thus it is necessary to design an implementation specific for the affine LoG construction.

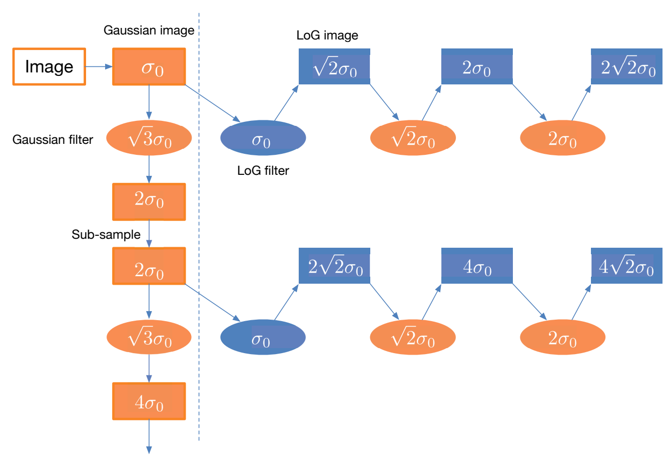


Fig. 3. The structure to implement the affine LoG.

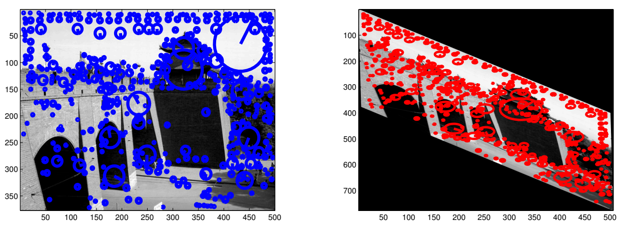
Figure 3 presents the implementation structure we propose for the affine LoG construction. It is based on the properties of affine Gaussian and affine LoG. This structure can be divided into two parts, Gaussian and LoG. The separation is illustrated by a blue dotted line in Figure 3. In this structure, the initial image in each octave is generated by a Gaussian pyramid, which is composed by a recursive smoothing and sampling operation. In this part, just one Gaussian filter is recursively applied with the scale $\sqrt{3}\sigma_0$ to create all the initial images for the LoG. Thanks to the sub-sampling property, this pyramid structure can help to generate the Gaussian blurred images with the scales $\sigma_0, 2\sigma_0, \cdots$. Three other filters, including two affine Gaussian filters and one LoG filer, are used to generate the rest of LoG images for each octave. Because of the cascade implementation, the LoG filter will be first applied, guaranteeing that all the other generated images are all LoG filtered. Thanks to the Gaussian pyramid, the whole LoG scale space can be generated as efficiently and accurately as the classical pyramid structure.

It should be noted that we cannot generate a Gaussian scale space in this pyramid at the same time. If it is also required to obtain a Gaussian scale space, which may be used for generating the gradient descriptor, the LoG filter shall be changed to a Gaussian filter to generate the corresponding Gaussian scale space, for which the structure only differs on the LoG filter. For a normal application, only 5 steps are required to generate the whole LoG scale space or Gaussian scale space by our proposed implementation structure. So if the application only requires an automatic extraction of features from the derivative of scale space, our proposed implementation structure will be very useful. If a Gaussian scale space is also required for some other applications, a change from a LoG filter to a Gaussian filter can easily provide this.

V. Experiments and Result

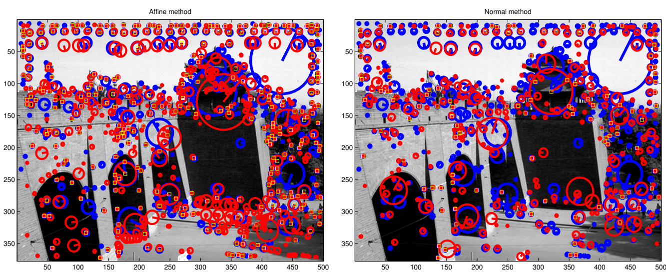
The performance of the affine scale space implementation can be evaluated by applying the feature detection to the images acquired under different view points. A typical feature detection on images with different view point is illustrated in Figure 4. In the figure, (a) shows the feature detection on the original image and (b) shows the feature detection on an affine transformed image, where the affine transform is employed to simulate the images with different view point. In figure (a), a classical scale space will be generated to detect the features from the ordinary view point and in figure (b), the detection will be based on an affine transformed scale space for an affine transformed view point. The performance of our proposed affine scale space implementation will be evaluated by comparing the extracted features from these two images respectively, given knowledge of the applied transformation.

In particular, we count how many of the detected features are still retained by employing the affine scale space to compensate for the affine transformation. To compare the detected features from the different perspectives of the image, the features detected from the affine transformed image are back-projected to the original one. In the figures 5, the blue circles depict the features detected from the original image,



(a) (b)

Fig. 4. Detection performance evaluation on the affine transformed images (a) Feature detection on the original image. (b) Feature detection on the affine transformed image.



(a) (b)

Fig. 5. Implementation performance evaluation by comparing feature detection (a) Feature detection by employing affine scale space. (b) Feature detection by classical scale space. The blue circles depict the features detected from the original image, red ones are the back-projected features detected from the affine transformed image. The radius of the circle represents its scale and the yellow box depict the correctly detected features. The correct detection by affine scale space is 0.6416, while for the conventional scale space it is only 0.2425.

red ones are the back-projected features detected from the affine transformed image. The radius of a circle represents the corresponding scale and the yellow box depicts the features that have been correctly detected by both the affine scale space and classical scale space. For the figures in Figure 5, the affine transformation is $\begin{bmatrix} 1.0000 & 0.0050 \\ 0.6000 & 1.0000 \end{bmatrix}$. In figure 5, the correct detection by affine scale space is 0.6416 but for the classical scale space it is just 0.2425.

Tilting	Citywall		Castle		Carton	
	affine	general	affine	general	affine	general
0.6	0.6416	0.2425	0.6541	0.2707	0.6133	0.2178
0.7	0.6202	0.2060	0.6016	0.1855	0.5711	0.1666
0.8	0.5923	0.1159	0.5841	0.1654	0.5224	0.1400
0.9	0.5494	0.1009	0.5689	0.1203	0.4659	0.1133
1.0	0.5773	0.0944	0.5689	0.1103	0.4756	0.1044
1.1	0.5516	0.0575	0.5213	0.0526	0.4867	0.0667
1.2	0.5343	0.0429	0.5414	0.0526	0.4579	0.0489

TABLE I

The correct detection rates for several other images are reported in the table I. As can be seen, the affine scale space consistently enables the correct match of a high number of keypoints, whereas the conventional system fails as most keypoints are not correctly matched after the affine transformation. Figure 6 also illustrates this correct matching ratios respectively by employing affine scale space and classical scale space. In this figure, we can see following the increase of view point difference, both the correct matching ratios of affine scale space and classical scale space will slightly decrease. However, the feature detection based on affine scale space always maintains a better performance than on the classical scale space specific to the view point invariant feature detection.

In our experiment, it takes 2.63 seconds to extract the features by employing the implementation we have proposed. The experiments are implemented on the computer with the CPU Inter Xeon(R) 5130 @ 2.00 GHz, RAM 4 GB and operating system 64 bit windows 7.

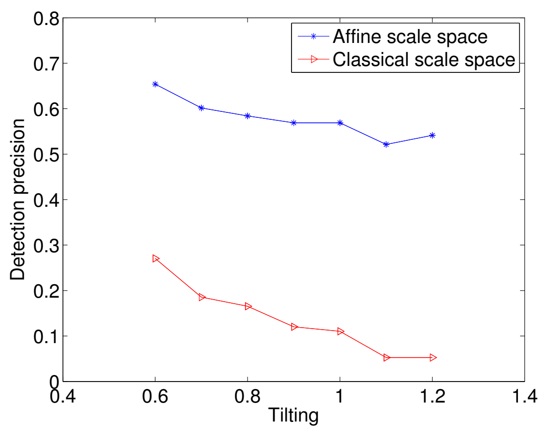


Fig. 6. Comparison of feature detection by employing affine scale space.

By this evaluation, it has been proved that the affine scale space can be successfully employed to improve the feature detection for the images under different view point, especially when the tilting angle is large.

Overall speaking, the affine scale space can guarantee that at least 30% of features will be retained without being affected by the affine transformation, whereas for the conventional scale space, only 1% of the features will be detected when the view point change is larger than 70°.

VI. Conclusion

The purpose of the affine scale space is to create a more general approach to the scale-invariant image representation providing invariance to affine transforms, by steering the Gaussian filters to the specific affine transformations.

In this paper, we have introduced the affine scale space to detect the features in an affine invariant way. We have also proposed a practical implementation structure, which is based on the properties of Gaussian axioms, to achieve real time operation. The affine scale space is a forward model, allowing to predict what will happen to an image under a different view point.

By the experiments, the affine scale space has been proved to have a better performance than the classical scale space especially when the difference of view point is large which can guarantee that at least 30% of features will be retained without being strongly affected by the change of view point.

References

[1] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys (CSUR)*, vol. 40, no. 2, p. 5, 2008.

[2] T. Lindeberg, *Scale-space theory in computer vision*. Springer Science & Business Media, 1993.

[3] ——, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus of attention," *International Journal of Computer Vision*, pp. 283–318, 1993.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] G. C. Stavros Paschalakis, Gianluca Francini, "Test model 12: Compact descriptors for visual search," MPEG, Strasbourg, France, testmodel ISO/IEC JTC1/SC29/WG11/N14961, 12 2014.

[6] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1615–1630, 2005.

[7] A. H. Salden, B. M. ter Haar Romeny, and M. A. Viergever, "Linear scale-space theory from physical principles," *Journal of Mathematical Imaging and Vision*, vol. 9, no. 2, pp. 103–139, 1998.