Abstract

This work presents a set of techniques concerning image forensics, more specifically, the core of the thesis deals with image clustering based on camera fingerprints. Image clustering using camera fingerprints is a blind problem, and clustering is performed in the absence of any prior information. The clustering process faces some serious issues of high computational cost, input-output (I/O) cost and $NC \gg SC$ problem, i.e., when the number of cameras NC is much larger than the average number of images per camera SC. Reducing the computational cost and finding a solution to the $NC \gg SC$ problem are the main objective of this thesis. Several algorithms have been proposed and presented, including reduced complexity image clustering (RCIC), fast image clustering algorithm based on fingerprints ordering (FICFO), canopy based image clustering (CIC) and compressed fingerprints based image clustering (CFIC) algorithms. The RCIC and FICFO are also implemented with an optional stage of attraction. The attraction process helps to improve the quality of the clusters. The RCIC and FICFO with attraction are represented as RCIC-A and FICFO-A, respectively.

For every algorithm, an estimate of the camera fingerprint for each available image is computed using standard techniques, then the algorithm tries to cluster those fingerprints according to the source camera. RCIC randomly selects a camera fingerprints from a set of un-clustered fingerprints as a reference and uses it as an attractor to construct a cluster. The remaining fingerprints are processed in the same manner by randomly choosing the next reference fingerprint out of the set of currently un-clustered fingerprints, and building another cluster. The RCIC is applied until all fingerprints are assigned to a cluster. The clusters are refined using the attraction process in the RCIC-A algorithm. The RCIC and RCIC-A clusters images with significantly lower computational cost in comparison with existing clustering algorithms, while maintaining similar or even better performance.

Moreover, these algorithms are robust to the $NC \gg SC$ problem. The clustering process is further simplified and made faster in FICFO by sorting the camera fingerprints using the inherent information of images. A factor of goodness called ranking index $\Re I$ is computed for each fingerprint, using the gray level, saturation, and textures level of the respective image. The higher the $\Re I$, the lower is the estimation error on the fingerprint and vice versa. Therefore, all the fingerprints are arranged in the descending order of the $\Re I$ and the best fingerprint i.e., the fingerprint on the top, among the un-clustered fingerprints, is selected as reference fingerprint to attract other fingerprints of the same camera and construct a cluster. The results confirm that the FICFO and FICFO-A result in a cluster of high quality and at lower computational complexity. The sorting of fingerprints helps to reduce the computational cost with respect to RCIC and RCIC-A. The method efficiently handles the problem of $NC \gg SC$. The results show that these algorithms are more suitable for large scale clustering.

The CIC algorithm is another approach to cluster images using the camera

fingerprints. The CIC algorithm uses the sorted camera fingerprints and operates in two stages. In the first stage, a relaxed threshold is set to construct fewer raw clusters of large sizes. The raw clusters are further clustered using a strict threshold. This results in a large number of pure clusters; most of them are singleton clusters. The clusters are refined using the attraction stage, which helps to reduce the number of clusters and improve the quality of clusters. The results show that the CIC algorithm does not suffer from the $NC \gg SC$ problem and has a significantly lower computational cost. However, even though the computational complexity of the CIC algorithm is lower than FICFO and RCIC, sometimes the performance of the CIC algorithm is degraded with respect to RCIC-A and FICFO-A. Furthermore, the lower computational complexity of the CIC algorithm for large image datasets makes it suitable for large scale clustering.

To reduce the computational cost further, reduced and full camera fingerprints are employed in the CFIC algorithm. The initial clustering is done using reduced fingerprints to construct clusters. The clustering performed on reduced fingerprints has a lower computational cost than that of using full fingerprints. The full fingerprints of each initially created cluster are merged by taking the average of them and standardizing the result to zero mean and unit norm. The merged fingerprints are used to refine the cluster and construct fine clusters. The CFIC algorithm, before clustering, computes $\Re I$ for each fingerprint, and arranges all fingerprints i.e., full and reduced fingerprints, in the descending order of $\Re I$. The CFIC method results in high quality clusters, comparable to the state-of-the-art techniques, at a significantly lower computational cost. The results show that the CFIC algorithm is suitable for large scale clustering and does not suffer from the $NC \gg SC$ problem.