GEOMETRIC TAMPERING ESTIMATION BY MEANS OF A SIFT-BASED FORENSIC ANALYSIS

Irene Amerini, Lamberto Ballan, Roberto Caldelli, Alberto Del Bimbo and Giuseppe Serra

MICC - Media Integration and Communication Center
University of Florence, Italy

ABSTRACT

In many application scenarios digital images play a basic role and often it is important to assess if their content is realistic or has been manipulated to mislead watcher's opinion. Image forensics tools provide answers to similar questions. This paper, in particular, focuses on the problem of detecting if a feigned image has been created by cloning an area of the image onto another zone to make a duplication or to cancel something awkward. The proposed method is based on SIFT features and allows both to understand which are the image points involved in the counterfeit attack and, furthermore, to recover the parameters of the geometric transformation. Experimental results are provided to witness the powerfulness of the proposed technique.

Index Terms— SIFT, image tampering, geometric transformation, image forensic, authenticity

1. INTRODUCTION

Looking at an image often raises a question, is it realistic or has it been retouched? [1]. Such a question is usually due to the well-known easiness with which digital images can be modified to alter their content and the meaning of what is represented in them. When the context in which pictures are used is not a tabloid or an advertising poster, but, for instance, it is a court of law where images are presented as basic evidences for a trial to influence the judgement, answering reliably to such questions about integrity becomes fundamental. Image forensics deal with these issues by developing technological instruments which generally allow to determine, only on the basis of a photograph, if that asset has been tampered with [2] or which has been the adopted acquisition device [3, 4]. Furthermore, it would be interesting, once established that something has happened, to understand what: if an object or a person has been covered, if a part of the image has been cloned, if something has been copied from another image or, even more, if a combination of these processes has been carried out. In particular, when an attacker creates his feigned image by cloning an area of the image onto another zone to make a duplication or to cancel something that was awkward, he is often obliged to apply a geometric transformation to satisfactorily achieve his aim. Succeeding in individuating if this kind of tampering has taken place and in estimating the parameters of the transformation occurred (i.e. horizontal and vertical translation, scaling factors, rotation angle) could be worthy in a forensic analysis. On the basis of such a consideration, in this paper a new methodology which answers to this requirement is presented. Such a technique is based on Scale Invariant Features Transform (SIFT) [5] algorithm which is used to robustly detect and describe clusters of points belonging to cloned areas. Successively, these points are exploited to reconstruct the parameters of the occurred geometric transformation.

The paper is structured as it follows: in Section 2 a brief description of SIFT technique is provided and in Section 3 the proposed method is discussed in detail; some experimental results, both to demonstrate forgery detection capability and to prove performances with regard to transformation parameters estimation, are debated in Section 4 and conclusions are drawn in Section 5.

2. SIFT FEATURES FOR IMAGE FORENSICS

Many techniques have been proposed to address the problem of copy-move forgery detection. Almost all methods divide the image into overlapping blocks and then a feature extraction process to represent the image blocks is performed. The forgery decision is made only if there are more than a certain number of blocks that are connected to each other and the distance between each duplicated block pair is the same. Bayram *et al*. [6] presented a comparative evaluation among Discrete Cosine Transform (DCT), Principal Component Analysis (PCA) and Fourier Mellin Transform (FMT) features. In particular, the robustness of these methods against rotation and scaling operation is reported. FMT and DCT methods can detect rotations of up to 10° and 5° respectively and they can not detect scaling over 10%, while the PCA method can not detect scaling and rotation transformation at all. Recently visual local features have become extremely popular for the tasks of object detection and recognition, due to their robustness with respect to partial occlusion, clutter and geometrical transformations. Many different approach have been presented but a common idea is to model a complex

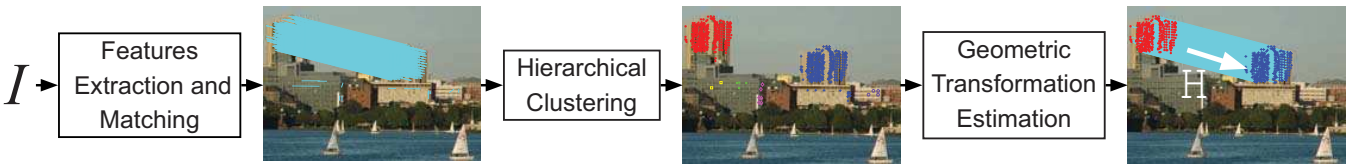


Fig. 1. Overview of our system. The lines link pairs of matched points and colors represent different clusters.

object or a scene by a collection of local salient points. These methods typically start with a detection step, in which interest points are localized, then representations of local patches are extracted by a descriptor, generally defined so as to be invariant with respect to orientation, scale and affine transformations. The review paper by Mikolajczyk and Schmid [7] provides a comprehensive analysis of several local descriptors, but perhaps the most commonly used solution is Scale Invariant Features Transform (SIFT) [5] because of their high performances and relatively low computational cost. Given an image, SIFT features are detected at different scales by using a scale space representation implemented as an image pyramid. The pyramid levels are obtained by Gaussian smoothing and sub-sampling of the image resolution while interest points are selected as local extrema (min/max) in the scale-space. These points (usually called *keypoints*) are extracted by applying a computable approximation of the Laplacian of Gaussian (LoG) following the same approach of the Hessian detector. In particular, the SIFT algorithm approximates LoG by iteratively computing the difference between two nearby scales in the scale-space. This idea is referred to as the Difference of Gaussians (DoG) approach. Once these keypoints are detected, SIFT descriptors are computed at their locations in both image plane and scale-space. Each descriptor consists in a histogram of 128 elements, obtained from a 16x16 pixels area around the corresponding keypoint. The contribution of each pixel is obtained by calculating image gradient magnitude and direction in scale-space and the histogram is computed as the local statistics of gradient directions (8 bins) in 4x4 sub-patches of the 16x16 area.

3. THE PROPOSED METHOD

Our method relies on SIFT features since they are, as previously introduced, robust to scaling, rotation and also to affine transformations. These properties are well-suited for the detection of forgeries in images. In fact, the copied part has the same appearance of the original one, thus keypoints extracted in that region will be quite similar to the originals. Therefore matching between SIFT features can be used to discover which part was copied and which geometrical transformation was applied. Figure 1 shows a schematization of the overall system. In particular, given an image I, we extract the keypoints $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and their SIFT descriptors $D = \{desc_1, \ldots, desc_n\}$. The best candidate match for each keypoint $\mathbf{x}_i$ is found by identifying its nearest neighbor from the other $n-1$ keypoints, which is the keypoint with the minimum Euclidean distance between their descriptors. In order to perform the matching decision (i.e. "are these two descriptors the same or not?"), evaluating the distance between two descriptors with respect to a global threshold does not perform well. This is due to the high-dimensionality of the feature space (128) in which some descriptors are much more discriminative than others [5]. We obtain a more effective measure by using the ratio between the distance of the closest neighbor to that of the second-closest one, and comparing it with a threshold. For sake of clarity, given a keypoint we define a similarity vector $S = \{d_1, d_2, \ldots, d_{n-1}\}$ that represents the sorted euclidean distances with respect to the other descriptors. The keypoint is matched only if the following constraint is satisfied: $\frac{d_1}{d_2} < T$ (fixed empirically to 0.6). Iterating on each keypoint in X, we can obtain the set of matched points.

Clustering. To detect the possible cloned areas we use agglomerative hierarchical clustering on the spatial location of the matched points. Hierarchical clustering creates a hierarchy of clusters which may be represented in a tree structure. The algorithm starts by assigning each point to a cluster; then it finds the closest (i.e. the most similar) pair of clusters and merges them into a single cluster. We consider the distance between two clusters to be equal to the shortest distance from any member of one cluster to any member of the other cluster. The final number of clusters is obtained by cutting the tree at a particular height. For the cutting criteria we use the inconsistency coefficient value. It characterizes each link in a cluster tree by comparing its length with the average length of other links at the same level of the hierarchy.

Geometric transformation estimation. Translation, rotation and scaling transformation between an original area and its copied area can be determined using the set of extracted matched points. Let the matched point coordinates be, for the two areas, $\mathbf{x}_i = (x, y, 1)^{\mathtt{T}}$ and $\mathbf{x}'_i = (x', y', 1)^{\mathtt{T}}$ respectively, their geometric relationships can be defined by an affine homography which can be represent by a 3×3 matrix as:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \mathtt{H} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{1}$$

This matrix can be computed by at least three matched points. In particular, we determine $\mathtt{H}$ by using Maximum Likelihood

estimation of the homography [8]. This method seeks homography $\mathtt{H}$ and pairs of perfectly matched points $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}'_i$ that minimize the total error function:

$$\sum_i \left[d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2 \right] \text{ subject to } \hat{\mathbf{x}}'_i = \mathtt{H}\hat{\mathbf{x}}_i \ \forall i. \quad (2)$$

However mismatched points (*outliers*) can severely disturb the estimated homography. The goal then is to determine a set of *inliers* from the presented correspondence so that the homography can be optimally estimated from these inliers using the algorithm above. For this purpose we apply the RANdom Sample Consensus algorithm (RANSAC) [9]. The geometric transformation can be so computed from an affine homography. In fact, $\mathtt{H}$ can be represented as:

$$\mathtt{H} = \begin{bmatrix} \mathtt{A} & \mathbf{t} \\ \mathbf{0}^{\mathtt{T}} & 1 \end{bmatrix} \quad (3)$$

where

$$\mathtt{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}. \quad (4)$$

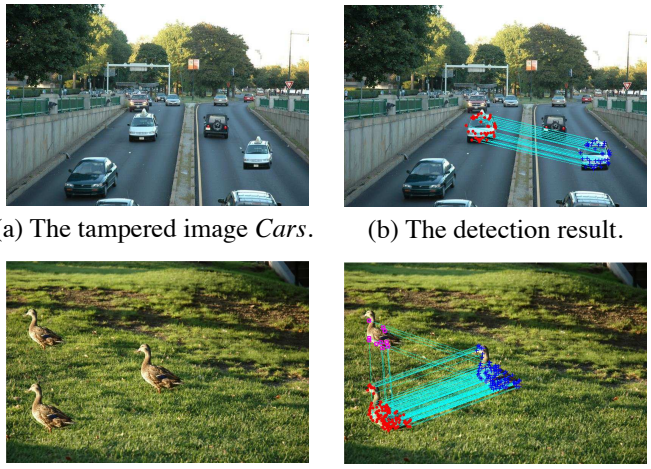
The vector $\mathbf{t}$ is the translation while the matrix $\mathtt{A}$ is the composition of rotation and non-isotropic scaling transformations. In particular, $\mathtt{A}$ can always be decomposed in SVD (Singular Value Decomposition) form as $\mathtt{A} = \mathtt{USV}^{\mathtt{T}}$, where $\mathtt{S} = \mathtt{diag}(s_1, s_2)$. Moreover, the matrix $\mathtt{A}$ can be also rewritten as: $\mathtt{A} = (\mathtt{UV}^{\mathtt{T}})(\mathtt{VSV}^{\mathtt{T}}) = \mathtt{R}(\theta)(\mathtt{R}(-\Phi)\mathtt{SR}(\Phi))$ since $\mathtt{U}$ and $\mathtt{V}$ are orthogonal matrices. Thus, $\mathtt{A}$ is considered to be the concatenation of a rotation Φ, obtained by the rotation matrix $\mathtt{R}(\Phi) = \mathtt{V}^{\mathtt{T}}$; a scaling $\mathtt{S}$, in which s_1 and s_2 respectively represents the (rotated) x and y directions; a rotation back (by $-\Phi$); and finally another rotation θ ($\mathtt{R}(\theta) = \mathtt{UV}^{\mathtt{T}}$).

4. EXPERIMENTAL RESULTS

The proposed approach has been tested on several images with different contents, coming from the Columbia photographic images dataset [10] and from a personal collection. In our experiments, we tampered these images by coping and pasting an image part (on the average 1.2% of the whole image) over another area in the same image. The dataset is composed by 100 tampered images obtained from 10 original images. These feigned images are obtained by making 10 different attacks by translating, rotating, scaling or both, the copied part of an image before pasting. Table 3 shows the geometric transformations of these attacks (from *a* to *j*). In particular for each attack is reported the rotation degree θ and the scaling factor s_x, s_y applied to the x or y axis of the original image part (e.g. in the attack *h*, the x axes is scaled by 40% and y by 20%). The value of the translation is not shown because each tampering requires a different translation, depending on the context of the image. Experimental results for copy-move forgeries detection and the analysis of the performances achieved in the geometric transformation estimation are presented in the following.

Attack	θ °	s_x	s_y	Attack	θ °	s_x	s_y
a	0	1	1	*f*	0	1.2	1.2
b	10	1	1	*g*	0	1.3	1.3
c	20	1	1	*h*	0	1.4	1.2
d	30	1	1	*i*	10	1.2	1.2
e	40	1	1	*j*	20	1.4	1.2

Table 3. The 10 different combinations of geometric transformations applied to the original patch.



(a) The tampered image *Cars*. (b) The detection result.

(c) The tampered image *Goslings*. (d) The detection result.

Fig. 2. On the left column the tampered images and on the right column the outputs of the proposed detection algorithm.

4.1. Forgeries detection

Each tampered image in our dataset is recognized as such, when at least three points, between the original and the cloned area, are found. In particular our method outperforms the others copy-move methods [6]; in fact, it is able to detect feigned images with cloned areas rotated above 10° and scaling above 20%. In the following, we give a detailed account to the results obtained with rotation value from 10° to 40° and scaling from 20% to 40%, but the proposed method is able to detect also rotation of 90° and scaling of 50%. A comprensive study on all the possible ranges of rotation and scaling will be investigated in the next future.

Two tampered images of the dataset are shown as representative examples on the left column of Figure 2, while on the right side the outputs of the detection algorithm for both images are pictured. So it is pointed out by Figure 2(b) "Cars" that we are dealing with a simple cloning, and from Figure 2(c) that we are dealing with a multiple cloning; in fact the gosling in the centre of the image was copied and then pasted, without manipulation, on the bottom left of the image and then downsized (20%) before it is pasted on the top left of the image.

A	t_x	$\hat{t}_x$	$\|e\|$	t_y	$\hat{t}_y$	$\|e\|$	θ	$\hat{\theta}$	$\|e\|$	s_x	$\hat{s}_x$	$\|e\|$	s_y	$\hat{s}_y$	$\|e\|$
a	304	304.02	0.02	80.5	81.01	0.51	0	0.040	0.040	1	1.004	0.004	1	0.998	0.002
b	304	305.20	1.20	80.5	82.42	1.92	10	9.963	0.037	1	1.001	0.001	1	0.999	0.001
c	304	305.55	1.55	80.5	82.64	2.14	20	20.009	0.009	1	1.006	0.006	1	0.998	0.002
d	304	305.04	1.04	80.5	82.49	1.99	30	30.092	0.092	1	1.002	0.002	1	0.998	0.002
e	304	306.08	2,08	80.5	78.43	2.07	40	39.932	0.067	1	1.007	0.007	1	1.004	0.004
f	304	304.88	0.88	80.5	80.41	0.09	0	0.080	0.080	1.2	1.202	0.002	1.2	1.198	0.002
g	304	305.07	1.07	80.5	79.87	0.63	0	0.108	0.108	1.3	1.304	0.004	1.3	1.303	0.003
h	304	305.78	1.78	80.5	80.18	0.32	0	0.037	0.037	1.4	1.403	0.003	1.2	1.206	0.006
i	304	305.23	1.23	80.5	81.76	1.26	10	9.910	0.090	1.2	1.203	0.003	1.2	1.201	0.001
j	304	305.02	1.02	80.5	80.82	0.32	20	20.067	0.067	1.4	1.404	0.004	1.2	1.198	0.002

Table 4. Transformation parameters estimation on image *Cars*. The values t_x and t_y are expressed in pixels while θ in degrees.

4.2. Transformation parameters estimation

Table 4 reports, for the image *Cars* and for each transformation parameter, the original value applied to the patch, the estimated one and the absolute error ($|e|$). In particular we observe that the estimated parameters are very close to the original values both for the rotation θ and the scaling values s_x and s_y. Also the translation vector of the duplicated parts is considered and the estimation errors between the original values t_x, t_y and the estimated $\hat{t}_x$, $\hat{t}_y$ are only few pixels. In Table 5, we report the Mean Absolute Error (MAE) values for all the attacks (summarized in Table 3) averaged on all the 10 images. The results show that our method returns in the worst case a maximum value of MAE, in case of translation attack, that is equal to 8.736 (in x) and 6.541 (in y). Moreover, rotation and scaling attacks returns very low MAE values (i.e. 0.513 for θ, 0.084 and 0.094 for scale).

A	MAE(t_x)	MAE(t_y)	MAE(θ)	MAE(s_x)	MAE(s_y)
a	1.069	4.481	0.120	0.002	0.002
b	8.097	6.149	0.513	0.013	0.014
c	1.707	6.328	0.220	0.011	0.007
d	5.397	5.534	0.200	0.010	0.010
e	8.736	6.541	0.081	0.012	0.005
f	4.014	5.557	0.255	0.084	0.094
g	1.584	3.892	0.092	0.045	0.039
h	1.471	3.902	0.151	0.014	0.008
i	2.313	5.160	0.122	0.005	0.006
j	3.624	5.601	0.269	0.081	0.018

Table 5. Transformation parameters estimation errors.

5. CONCLUSIONS

A new technique for image forensics based on SIFT features has been introduced. Its powerfulness to detect copy-move attack and to trace back the geometric transformation occurred has been witnessed by specific experimental results. Future works will be dedicated to investigate and improve the behavior of the technique in relation with the size and the texture of the cloned image patch.

Acknowledgements. This work is partially supported by the EU ICT 3D-COFORM Project (Contract FP7-231809) and IM3I Project (Contract FP7-222267).

6. REFERENCES

[1] S. Lyu and H. Farid, "How realistic is photorealistic?," *IEEE Trans. on Signal Processing*, vol. 53, no. 2, pp. 845–850, 2005.

[2] H. Farid, "A survey of image forgery detection," *IEEE Signal Processing Magazine*, vol. 2, no. 26, pp. 16–25, 2009.

[3] A. Swaminathan, M. Wu, and K.J.R. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 1, pp. 101–117, 2008.

[4] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining image origin and integrity using sensor noise," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 1, pp. 74–90, 2008.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[6] S. Bayram, H.T. Sencar, and N. Memon, "An efficient and robust method for detecting copy-move forgery," in *Proc. IEEE ICASSP*, 2009.

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

[8] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd edition, 2004.

[9] M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[10] T.-T Ng, S.-F. Chang, J. Hsu, and M. Pepeljugoski, "Columbia photographic images and photorealistic computer graphics dataset," Tech. Rep., ADVENT, Columbia University, 2004.