Copy-Move Forgery Detection and Localization by Means of Robust Clustering with J-Linkage

Irene Amerini[a], Lamberto Ballan[a,*], Roberto Caldelli[a], Alberto Del Bimbo[a], Luca Del Tongo[a], Giuseppe Serra[a,b]

[a]*Media Integration and Communication Center (MICC), Università degli Studi di Firenze, Viale Morgagni 65, 50134 Firenze, Italy*
[b]*Dipartimento di Ingegneria "Enzo Ferrari", Università degli Studi di Modena e Reggio Emilia, Via Vignolese 905/b, 41125 Modena, Italy*

Abstract

Understanding if a digital image is authentic or not, is a key purpose of image forensics. There are several different tampering attacks but, surely, one of the most common and immediate one is copy-move. A recent and effective approach for detecting copy-move forgeries is to use local visual features such as SIFT. In this kind of methods, SIFT matching is often followed by a clustering procedure to group keypoints that are spatially close. Often, this procedure could be unsatisfactory, in particular in those cases in which the copied patch contains pixels that are spatially very distant among them, and when the pasted area is near to the original source. In such cases, a better estimation of the cloned area is necessary in order to obtain an accurate forgery localization. In this paper a novel approach is presented for copy-move forgery detection and localization based on the J-Linkage algorithm, which performs a robust clustering in the space of the geometric transformation. Experimental results, carried out on different datasets, show that the proposed method outperforms other similar state-of-the-art techniques both in terms of copy-move forgery detection reliability and of precision in the manipulated patch localization.

Keywords: Digital image forensics, Tampering detection, Copy-move detection, Copy-move localization

1. Introduction

Nowadays, digital crime is growing at a rate that far surpasses defensive measures. Sometimes a digital media content, such as an image or a video, may be found to be incontrovertible evidence of a crime or of a malevolent action. By looking at a digital data as a digital clue, *multimedia forensics* technologies are introducing a novel methodology for supporting clue analysis and providing an aid for making a decision on a crime [1, 2]. Multimedia forensics deals with developing technological instruments which generally allow to determine, without any additional information inside the image (e.g. a watermark), if that asset has been tampered with or which has been the adopted acquisition device. In particular, tampering detection refers to the problem of assessing the authenticity of digital images [3], and this is the topic of this paper.

Information integrity is fundamental in a trial, but it is clear that the advent of digital pictures and relative ease of digital image processing makes today this authenticity uncertain. An example of this problem, that recently appeared in a Tunisian newspaper, is given in Figure 1; here the photo has been tampered with in order to make the crowd appear larger. It demonstrates that this kind of manipulation is used more and more often in news and advertising campaigns. Modifying an image to change the meaning of what is represented in it could be crucial when this digital data is used in a court of law, where it can be presented as basic evidence to influence the judgment. Furthermore, in case of tampering, it would be interesting to understand what kind of manipulation has been applied: for example 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, and so on.

In this paper we address this issue, in particular by detecting if a *copy-move attack* has taken place (i.e. when the attacker creates his feigned image by cloning an area of the image onto another zone) and by localizing the tampered area in the image. The proposed method relies on Scale Invariant Features Transform (SIFT) [4]

*Corresponding author. Tel.: +39 055 4237409.
Email addresses: irene.amerini@unifi.it (Irene Amerini), lamberto.ballan@unifi.it (Lamberto Ballan), roberto.caldelli@unifi.it (Roberto Caldelli), alberto.delbimbo@unifi.it (Alberto Del Bimbo), lukadt@gmail.com (Luca Del Tongo), giuseppe.serra@unimore.it (Giuseppe Serra)

Figure 1: The figure reports the photo published on the front page of Le Maghreb, a tunisian newspaper, on January 2012. The photo was digitally altered duplicating the crowd to appear larger.

and features matching, and improves our previous work [5, 6] by introducing a new robust clustering phase based on the J-Linkage algorithm [7], and an accurate forgery localization procedure. The localization of the duplicated region has been set up on the basis of the clusters obtained in the previous phase. This is done using ZNCC (Zero mean Normalized Cross-Correlation) between the original image and the warped image obtained from the estimated geometric transformation occurred in the tampering attack. In order to obtain an accurate localization, it is necessary to have an effective clustering procedure (like the one presented in this paper) that is able to guarantee a good estimate of the geometric transformation.

The rest of this paper is organized as follows. In Section 2, we discuss the existing works concerning the detection of copy-move forgeries. Section 3 presents the proposed copy-move forgery detection and localization method, while Section 4 contains experimental results. Conclusions are finally drawn in Section 5.

2. Related works

As discussed earlier, copy-move manipulations involve concealing or duplicating one region in an image by overlaying portions of the same image on it. In order to address the problem, researchers have developed various techniques which can be classified into two main categories: *block-based* and visual *feature-based* methods.

2.1. Block-based methods

These methods seek a dependence between the image original area and the pasted one, by dividing the image into overlapping blocks and then applying a feature extraction process in order to represent the image blocks through a low dimensional representation. Different block-based representations have been previously proposed in the literature, such as Principal Component Analysis (PCA) [8, 9], Discrete Cosine Transform (DCT) [10] and Discrete Wavelet Transform (DWT) [11, 12], for both tasks of copy-move detection [10, 11, 13, 9] and image splicing [14]. Recently, in the study of Bashar *et al.* [15], the authors proposed a duplication detection approach that can adopt two robust features based on DWT and kernel principal component analysis (kPCA). A different kind of features are used in [16], in fact the authors choose the averages of red, green and blue components with other four features, computed on overlapping blocks, obtained by calculating the energy distribution of luminance along four different directions. To improve the computational complexity of these methods, in [17] the authors proposed to use the radix sort for sorting the feature vectors of the divided sub-blocks, as an alternative to lexicographic sorting, which is commonly adopted. However, all these methods assume that the copied region has not undergone any post-processing such as scaling, rotation and JPEG compression.

To deal with this issue, a preliminary work by Mahdian *et al.* has been presented in [18] where the authors proposed a block-based representation calculated using blur invariants. They used PCA to reduce the number of features and a k-tree to identify the interested regions. Authors in [19] proposed a different kind of feature that is based on the Fourier-Mellin Transform that is invariant to small rotation and resizing of the copied regions. However, the technique fails when the rotation and the resizing is significant. This method was improved in [20] in which better rotation invariance was achieved by taking projections along angular directions instead of radius direction. However, also in this case the scale invariance seems to be valid only over a small range, and the number of false positives yielded is quite high.

Recently, methods more robust to reflection, rotation and scaling have been proposed in the literature. In [21] overlapping blocks of pixels are mapped into log-polar coordinates, and then summed along the angle axis, to obtain a one-dimensional descriptor invariant to reflection and rotation. Wang et al. [22] proposed the use of circle region instead of square block and adopt as feature the mean of the intensities of the circle region with different radii to overcome the effect of rotation. Ryu *et al.* [23] exploited the Zernike moments as features since their magnitude is algebraically invariant to rotation transformation. To this end, a more general approach is presented in [24], in which is reported a technique to better detect variations in rotation and scaling in the copied part by introducing a post-processing

phase for the block selection, instead of the widely-used shift vectors. The authors called this stage Same Affine Transformation Selection (SATS) and it is collocated after the feature extraction and block matching phases. In particular, they show that any set of rotation-invariant features like [21, 22, 23] can benefit from the inclusion of this processing step in the pipeline.

2.2. Visual feature-based methods

It has been demonstrated that block based methods often result in significant false positives. Moreover, invariance to geometrical transformations and to other manipulations like flipping, brightness changes and blurring is hard to establish [24]. Feature-based techniques try to avoid these problems by choosing to match features in the image, instead of blocks, using local visual features like Scale Invariant Feature Transform (SIFT) or Speed Up Robust Features (SURF). In particular, these features have been widely used for image retrieval and object recognition due to their robustness to several geometrical transformations (e.g. rotation, scaling and affine transformation).

Some works have recently appeared on copy-move forgery detection based on SIFT [25, 6] or SURF features [26]. In the work of Pan and Lyu [25], SIFT features are chosen in order to localize the copied region through the use of a correlation map. However, quantitative results on a realistic dataset are not given and the method does not consider the case of multiple forgeries accurately. Multiple copy-move forgeries are instead managed in [6] by performing a robust SIFT feature matching procedure and then a clustering of the keypoints coordinates in order to separate the different cloned areas. Anyway, the method is used only for copy-move detection and not for accurate tampering localization. Kakar and Sudha [27] proposed to use MPEG-7 features in order to detect and localize copy-move forgeries, by following a very similar framework to [6].

Although methods such as [6, 27] have demonstrated good performance in copy-move detection, sometimes clustering and localization could be unsatisfactory. In particular in those cases in which the copied patch contains pixels that are spatially very distant among them, and when the pasted area is near to the source. In such cases, a better estimation of the cloned area is necessary in order to obtain an accurate forgery localization. In this paper we address this problem and we present a novel approach based on an adaptation of the J-Linkage algorithm.

3. The proposed method

We present a novel approach for detecting copy-move forgeries based on SIFT features and J-Linkage clustering. A schema of the whole system is shown in Figure 2. The first step consists of SIFT feature extraction and keypoint matching, the second step is devoted to the clustering and forgery detection, while the third one localizes the copied region, if a tampering has been detected. We summarize the whole procedure for tampering detection and localization in Algorithm 1.

3.1. Feature extraction and keypoint matching

The first step in our approach is based on SIFT features since they are robust to scaling, rotation and affine transformations that are well-suited for the detection of copy-move forgeries as has been recently demonstrated in [25, 6]. We detect keypoints that are stable local extrema in the scale space and, for each of them, a feature vector is computed from a local pixel area around the detected point. Given a test image I, let $\mathcal{S} := \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ be the list of n interest points taken from this image, where $\mathbf{s}_i = \{\mathbf{x}_i, \mathbf{f}_i\}$ is a vector containing the keypoint coordinates $\mathbf{x}_i = (x, y)$ and $\mathbf{f}_i$ is the feature descriptor of the local patch around the keypoint (i.e. an histogram of gradient orientations of 128 elements).

In presence of a copy-move manipulation the extracted SIFT keypoints from the copied and the original regions have similar descriptor vectors. Therefore, matching among SIFT features is adopted to detect if an image has been tampered with and, subsequently, localize such forgery. The simplest approach to match keypoints is to fix a global threshold on the Euclidean distance between descriptors but, due to the high-dimensionality of the feature space, this approach obtains a low accuracy because some descriptors are much more discriminative than others. For this reason Lowe [4] considers, given a keypoint, not only the distance with the first most similar keypoint, but also with the second one; in particular, he uses the ratio between the distance to the candidate match and the distance to the second similar feature point (i.e. the so-called 2NN test). To declare a match, this ratio must be lower than a fixed threshold τ (often equal to 0.6). This technique works well when a region is copied one time, but not if it is copied several times. To deal with this case, we use a generalization of Lowe's matching technique (g2NN test) recently proposed by Amerini *et al.* [6].

The g2NN starts from the observation that in a high-dimensional feature space such as that of SIFT features, keypoints that are different from the one considered

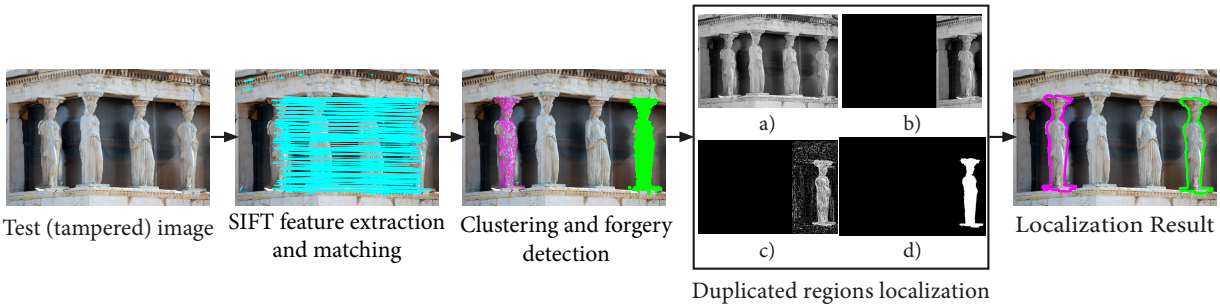

Figure 2: An outline of the proposed framework. The framework is composed by three steps: the first step consists of SIFT feature extraction and keypoint matching, the second step performs clustering and forgery detection, the third step localizes the forgery.

share very high and very similar values (in terms of Euclidean distances) among them. Instead, similar features show low Euclidean distances respect to the others. The idea of the 2NN test is that the ratio between the distance of the candidate match and the distance of the 2^{nd} nearest neighbor is low in the case of a match (e.g. lower than 0.6) and very high in case of two "random features" (e.g. greater than 0.6). Our generalization consists in iterating the 2NN test between d_i/d_{i+1} until this ratio is greater than τ (in our experiments we set this value to 0.5). If k is the value in which the procedure stops, each keypoint in correspondence to a distance in $\{d_1, \ldots, d_k\}$ (*where* $1 \leqslant k < n$) is considered as a match for the inspected keypoint. Using this g2NN strategy on all the keypoints S, we obtain a set of q matched pairs $\mathcal{P} := \{\mathbf{p}_1, \ldots, \mathbf{p}_q\}$, where $\mathbf{p}_i = (\mathbf{s}, \mathbf{s}')$. It allows, in the following steps, to identify the duplicated regions and therefore detect if the image has been tampered with.

3.2. The J-Linkage clustering and our copy-move detection strategy

A way to detect possible duplicated regions is to use a clustering algorithm on the coordinates of the keypoints, such as a hierarchical agglomerative clustering (HAC) procedure as in [6]. Following this approach, the clustering is performed by taking into account only the coordinates of the matched pairs and not the matching constraint between points. This method, like all the others clustering on spatial location, has two main drawbacks: *i)* the inability to separate duplicated regions that are close to each other, and *ii)* the difficulty to identify a patch as single, when it contains keypoints with a non-uniform spatial distribution (see Figure 3c). For these reasons, we proposed to design a clustering technique that does not work in the spatial domain of matched points but in the transformation domain. In particular, we introduce an adaptation of the J-Linkage algorithm [7] that is able to solve the aforementioned main drawbacks of a spatial clustering procedure (see Figure 3d).

The clustering starts with a random sampling on matched pairs, in order to generate m affine transformation hypotheses. For each pair, a preference set vector (PS) is defined indicating which transformations the pair prefers. Formally, given a matched pair $\mathbf{p}$ and m transformations $\mathcal{T} := \{T_1, \ldots, T_m\}$, the preference set vector $PS(\mathbf{p})$ is defined as $\{PS_1(\mathbf{p}), \ldots, PS_m(\mathbf{p})\}$, in which $PS_i(\mathbf{p})$ is defined as following:

$$PS_i(\mathbf{p}) = \begin{cases} 1 & \text{if } \mathbf{p} \text{ is an inlier of } T_i, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

It means that the distance between the model T_i and the matched pair $\mathbf{p}$ is less than a fixed threshold. In this way each pair is represented in a *conceptual space* $\{0, 1\}^m$. Since the matched pairs between the original and the duplicated regions share similar transformations, they will have similar conceptual representations.

The preference set vectors are then used in a hierarchical agglomerative clustering in order to find the transformations between the original and the cloned areas. This clustering algorithm starts by assigning each preference set vector to a cluster; then, for each step of the algorithm, the two clusters with smallest distance in the conceptual space are merged. The preference set vector of a cluster is computed as the intersection of the preference sets of matched pairs, and the distance between two clusters is computed as the Jaccard distance (J_δ) between the respective preference sets. Given two sets A and B, the Jaccard distance is defined as

$$J_\delta(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}; \tag{2}$$

this distance measures the overlapping degree of two sets. Identical sets have distance equal to 0, while disjoint sets have distance 1. According to this distance,

(a) (b) (c) (d)

Figure 3: a) Original images; b) Tampered images: purple lines depict the cloned regions; c) Clustering results with HAC; d) Clustering results obtained with the proposed method.

the cut-off value of the clustering is set to 1, which means that elements are merged until their preference sets overlap (or more intuitively until matched pairs share the same transformation). As a result of this procedure, each cluster will have at least one transformation shared by all its matched pairs. If more transformations fit with all the elements contained in the cluster, they should be very similar; therefore the final transformation is estimated by least squares fitting. In our algorithm, all the transformations that fit with a number of elements less than a fixed threshold N are discarded in order to remove possible outliers; this aspect has been further investigated in the experimental section in which different detection results are given changing this N value. Finally, if one transformation (or more) is detected, our system declares that the image has been altered by a copy-move attack.

Two aspects need to be clarified: *how to sample the matched pairs* for the transformation estimation and *how to compute a geometric transformation hypothesis*.

3.2.1. Sampling strategy

The strategy used to select a minimal sample set of matched points in [7], i.e. the original J-Linkage implementation, is based on the method of Kanazawa *et al.* [28]. It randomly selects an initial pair $\mathbf{p} = (\mathbf{s}, \mathbf{s}')$, from all the pairs $\mathcal{P}$, and it chooses from the remaining correspondences by fixing a high probability in the proximity of the first point $\mathbf{s}$. More precisely, let $\mathbf{x}_i$ be the coordinate of the keypoint $\mathbf{s}$, a new point $\mathbf{x}_j$ is selected with the following probability:

$$P(\mathbf{x}_j|\mathbf{x}_i) = \begin{cases} 0 & \text{if } \mathbf{x}_j = \mathbf{x}_i \\ \frac{1}{Z} \exp^{-\frac{\|\mathbf{x}_j - \mathbf{x}_i\|^2}{\sigma^2}} & \text{if } \mathbf{x}_j \neq \mathbf{x}_i \end{cases} \quad (3)$$

where Z is a normalization constant and σ is chosen heuristically (in [7] this parameter is set to 0.2). Following this procedure, the final set of points depends significantly on the parameter σ. This fact may result to a choice of the points that are too close or too far among them, leading to a rough estimation of the transformation. Moreover, this strategy is not able to deal with multiple cloned regions.

For these reasons, we have introduced a novel selection strategy (an example of this method is shown in Figure 4). Firstly, as in the previous method, we randomly select a matched pair $\mathbf{p} = (\mathbf{s}, \mathbf{s}')$ from the set $\mathcal{P}$. Then, we define two sets $\mathcal{O}$ and $\mathcal{D}$ of w nearest neighbors to the keypoint coordinates of $\mathbf{s}$ and $\mathbf{s}'$, respectively. In our experiments we fixed the parameter w to 12. The other k pairs that are necessary to find a minimal sample set for the transformation estimation ($k = 2$ in the case of an affine homography), are obtained randomly by selecting pairs from $\mathcal{P} := \{(\mathbf{s}, \mathbf{s}')_1, \ldots, (\mathbf{s}, \mathbf{s}')_q\}$, such that $\mathbf{s} \in \mathcal{O}$ and $\mathbf{s}' \in \mathcal{D}$.

As previously mentioned, the proposed strategy is able to handle multiple cloned regions. For example, if an original area is copied two times (e.g. $\mathcal{D}$ and $\mathcal{D}_1$ in Figure 4) a SIFT point in $\mathcal{O}$ would match respectively with a SIFT point in $\mathcal{D}$ and in $\mathcal{D}_1$. On the other hand in such a case, the strategy proposed by Kanazawa *et al.* should choose points by only considering the spatial proximity to the first point s. In this way, it would hap-

Algorithm 1: Tampering detection and localization

Input: A test image I; parameters: τ, m, w, N.
Output: A boolean value determining whether the test image has been tampered; the estimate (at the pixel level) of the cloned areas.

Extract SIFT from I, and let $\mathcal{S} := \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ be the list of interest points taken from this test image;

`// SIFT matching using our g2NN test`
for $i \leftarrow 1$ **to** n **do**
  for $j \leftarrow 1$ **to** n **do**
    Starting from points in $\mathcal{S}$, compute the set $\mathcal{P}$ of the q matched pairs using τ as threshold for the g2NN test;

`// Generate` m `affine transformation hypotheses`
for $i \leftarrow 1$ **to** m **do**
  Select a random pair $\mathbf{p} = (\mathbf{s}, \mathbf{s}')$ from $\mathcal{P}$;
  Define the sets $\mathcal{D}$ and $\mathcal{O}$ of w neighbors to the keypoint coordinates of $\mathbf{s}$ and $\mathbf{s}'$, respectively;
  Select two other random pairs $(\mathbf{s}, \mathbf{s}')$ such that $\mathbf{s} \in \mathcal{O}$ and $\mathbf{s}' \in \mathcal{D}$ (see Section 3.2.1);
  Use these three pairs to generate the affine transformation hypothesis T_i;

for $j \leftarrow 1$ **to** q **do**
  For each matched pair $\mathbf{p}_j$ in $\mathcal{P}$, compute the preference set vector $PS(\mathbf{p}_j)$;

Run Hierarchical Agglomerative Clustering on all the $PS(\mathbf{p})$ to find the set $\mathcal{T}$ of the transformations between the original and the cloned areas;
Define $\mathcal{T}'$ as the set obtained by removing all the transformations in $\mathcal{T}$ with less then N elements;
if *number of transformations in* $\mathcal{T}' > 0$ **then**
  Compute the localization of the duplicated regions for each $T_i \in \mathcal{T}'$ (using ZNCC);
  return *true*;
else
  return *false*;

pen that a pair $\mathbf{r}_1$ is formed by a point $\mathbf{z}$ close to $\mathbf{s}$ in $\mathcal{O}$, but its corresponding point $\mathbf{z_1}$ is far from $\mathbf{s}'$. This may lead to an inaccurate estimation of the homography between the regions $\mathcal{O}$ and $\mathcal{D}$. Instead our method is able to choose a pair by considering the proximity both to $\mathbf{s}$ and $\mathbf{s}'$, for example the pair $\mathbf{r}'$. Following the same procedure, the method is able to accurately estimate also the homography between the regions $\mathcal{O}$ and $\mathcal{D}_1$.

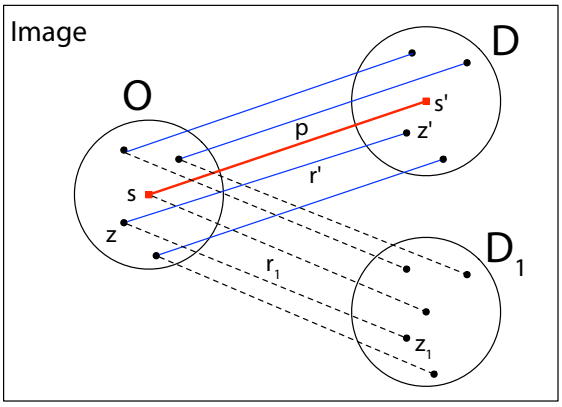

Figure 4: Representation of our sampling strategy.

3.2.2. *Computing a geometric transformation hypothesis*

The coordinates of the matched pairs previously selected, $\{(\mathbf{x}, \mathbf{x}')_1, \ldots, (\mathbf{x}, \mathbf{x}')_{k+1}\}$, are used to estimate the geometric transformation hypothesis. In particular, we use affine transformations in order to model the geometric distortions between the original and the copied regions (such as scaling, rotation, and shearing).

Formally, this kind of transformation can be expressed in matrix form as:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \mathbf{H} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

where a_{11}, a_{12}, a_{21} and a_{22} encode the rotation and scaling directions deformation, while t_x and t_y are the translation factors. An affine transformation has six degrees of freedom, corresponding to the six matrix elements, then the transformation can be computed from three matched pairs that are not collinear. In particular, to compute this estimation, we use the normalized Direct Linear Transformation (DLT) algorithm for affine homography (see Hartley and Zisserman [29]). So, given a set of correspondences $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{(k+1)})$ and $(\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_{(k+1)})$, the algorithm minimizes the following objective function:

$$\sum_{i=1}^{k+1} \|\mathbf{x}'_i - \mathbf{H}\mathbf{x}_i\|^2. \tag{4}$$

This linear method allows to quickly determine the m affine transformation hypotheses instead of a non-linear algorithm. In fact, in order to generate a good transformation hypothesis, m should be high due to the random sampling of the pairs (in our experiments $m = 500$).

3.3. Localizing duplicated regions

If an image is detected as a forgery, our system allows to obtain an accurate localization of the duplicated regions. This is another improvement with respect to our previous method [6]. The basic idea of our localization approach is that, given the estimated transformation between two sets of matched pairs, we can extend this transformation to the underlying dense regions in which it was really done. In fact, all pixels of the original region $\mathbf{R}^O$ are related to the pixels of a duplicated region $\mathbf{R}^D$, through the same transformation T (expressed in matrix form as $\mathbf{H}$):

$$\mathbf{R}^D = \mathbf{H}\mathbf{R}^O, \quad \mathbf{R}^O = \mathbf{H}^{-1}\mathbf{R}^D. \tag{5}$$

Applying the estimated transformation on the entire image, we will obtain a warped image in which the region $\mathbf{R}^O$ will overlap the region $\mathbf{R}^D$ (see Figure 2). In the same way, applying the inverse transformation $\mathbf{H}^{-1}$, the region $\mathbf{R}^D$ will overlap the region $\mathbf{R}^O$.

In order to localize the duplicated regions, we use a block-wise correlation measure based on Zero mean Normalized Cross-Correlation (ZNCC) between the gray-scale of the original image I (Figure 2a) and the warped image W (Figure 2b). It is computed as

$$\mathrm{ZNCC}(\mathbf{x}) = \frac{\sum_{\mathbf{v}\in\Omega(\mathbf{x})}\left(I(\mathbf{v})-\bar{I}\right)\left(W(\mathbf{v})-\bar{W}\right)}{\sqrt{\sum_{\mathbf{v}\in\Omega(\mathbf{x})}\left(I(\mathbf{v})-\bar{I}\right)^2\left(W(\mathbf{v})-\bar{W}\right)^2}}, \tag{6}$$

where $\Omega(\mathbf{x})$ is a 7 pixels neighboring area centered at every pixel $\mathbf{x}$ of I; $I(\mathbf{v})$ and $W(\mathbf{v})$ denote the pixel intensities at the location $\mathbf{v}$; $\bar{I}$ and $\bar{W}$ are the average pixel intensities of I and W, computed on $\Omega(\mathbf{x})$. Once the correlation map is obtained, we apply a Gaussian filter of size 7 pixels with standard deviation 0.5 in order to reduce the noise (Figure 2c). Next a binary image is created by thresholding the correlation map ($th = 0.55$). We discard all the small isolated regions that have an area less then 0.05%. Finally, mathematical morphological operations are used to fill eventually holes in the binary image (Figure 2d).

4. Experiments

We have evaluated the performance of the proposed method both from the point of view of forgery detection capability (*authenticity detection*) and also for what concerns the ability to recover copied areas (*patch localization*). Firstly, the proposed method is set up and improved by means of the design of a novel sample set selection strategy (subsection 4.3). Here we provide also evidence about the effectiveness of our approach, both in terms of detection and localization, on a novel realistic dataset called MICC-F600. Successively, we present a comparison with our previous work [6] to evaluate the improvement in terms of reliability in image forgery detection (subsection 4.4). Finally, a comparison regarding the capacity to localize manipulated patches is carried out towards other state-of-the-art techniques (subsection 4.5).

It has to be noticed that our method is also very competitive in terms of computational time. In contrast to other popular methods, our approach is able to effectively detect and localize copy-move forgery in a full-resolution image (e.g. 800×600 pixels) in around $8s$ on an Intel Q6600 with 4-GB RAM. This is a key aspect since several techniques are not used in real applications because of their computational complexity. For instance, block based methods are very expensive since features needs to be extracted from millions of blocks and they need to be sorted out. As an example, two of the most popular approaches of this kind of techniques [10, 8] are able to process the same images in around $295s$ and $71s$, respectively, according to our previous experiments on processing time requirements [6].

4.1. Data collections

Experimental results are reported on three different datasets: MICC-F2000 [6], SATS-130 [24], MICC-F600. For each of these, an example image is shown in Figure 5.

The MICC-F2000 dataset, introduced in [6], is composed of images with disparate contents coming from the Columbia photography image repository [30] and from a personal collection. Such a dataset consists of 2000 photos of 2048×1536 pixels: 1300 are original while 700 are tampered. The tampered images are obtained by applying 14 attacks such as translation, rotation, scaling, or a combination of them. The duplicated patches (corresponding to an average size of 1.12% of the whole image) are rectangular and they have not been accurately segmented and spatially well separated from the original areas (see for example Figure 5, on the top).

The second dataset is the SATS-130, adopted in [24] and made available by the authors. It is composed by 130 tampered images of different resolutions (from 420×300 to 3888×2592 pixels) and it does not contain not-tampered photos, so it can not be used for image authenticity detection tests. Forged images are obtained starting from 10 original images in which the copied regions are rotated by an angle which ranges from 0^o (pure translation) to 180^o, with steps of 15^o. An example for this dataset is reported in Figure 5 (center).

The patch localization binary masks are available and we have used this dataset as a benchmark for the evaluation of the proposed method in patch localization with respect to other state-of-the-art methodologies.

Finally, we have introduced a novel dataset named MICC-F600, containing realistic and challenging tampering attacks (Figure 5, on the bottom). It has been derived from the first two datasets, from other images provided by the authors of SATS-130 (Riess *et al.* [24])), and other images with multiple copied regions produced by ourselves. The images have different resolutions ranging from 800×533 to 3888×2592 pixels. MICC-F600 has been constructed with the aim to generalize, as much as possible, the kind of images under analysis. It consists of 600 images: 448 original, taken randomly within the 1300 original of MICC-F2000, and 152 forged, created starting from 38 non-tampered images (10 taken from SATS-130 and 28 new ones) in the following manner:

- 38 images in which a region is duplicated once through a translation;
- 38 images in which a region is duplicated twice or three times through a translation;
- 38 images in which the copied region is rotated by 30^o;
- 38 images in which the copied region is rotated by 30^o and scaled by 120%.

The MICC-F600 dataset is very challenging, it does not contain only forged images and, furthermore, the fake ones have not been created automatically and the tampering regions have different sizes and shapes. The manipulated patches have been cut out and post-processed to fit well in a realistic fashion within the area in which they have been pasted. Duplicated regions are not always spatially well separated (e.g. Figure 8d), their shapes can be quite challenging such as a fountain or a tree (e.g. Figures 8b, 8f). Moreover, several images contain multiple copy-move cloning (e.g. Figures 8a, 8e, 8h).

4.2. Evaluation criteria

To carry out performance evaluations, two set of metrics have been considered. In the case of *authenticity detection*, True Positive Rate (TPR) and False Positive Rate (FPR) have been computed; TPR is the fraction of tampered images correctly identified as such, while FPR is the fraction of original images that are not correctly identified:

$$\text{TPR} = \frac{\text{\# images detected as forged being forged}}{\text{\# forged images}},$$

$$\text{FPR} = \frac{\text{\# images detected as forged being original}}{\text{\# original images}}.$$

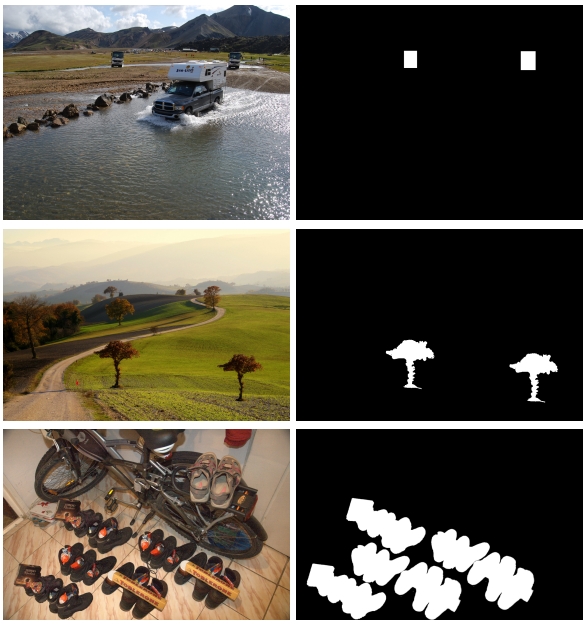

Figure 5: Example images of the adopted datasets: MICC-F2000 (top), SATS-130 (center) and MICC-F600 (bottom). The second column reports the ground-truth masks.

It is worth to state that authenticity detection is here intended by referring to the whole image, not only to the tampered patches. Thus, an image is labeled as forged if at least an affine transformation is estimated between a couple of image regions.

On the other side, the performance on *patch localization* is computed as the percentage of erroneously matched pixels F_P (i.e. false positives) and erroneously missed pixels F_N (i.e. false negatives). Formally, let $\mathbf{R_1}$ be the copied region, $\mathbf{R_i}$ ($i > 1$) be the i^{th} duplicated region, and $\mathbf{B}$ the unchanged background; then F_P and F_N are defined as

$$F_P = \frac{|\text{matches in } \mathbf{B}|}{|\mathbf{B}|} \tag{7}$$

and

$$F_N = \frac{|\text{missed matches in } (\bigcup_i \mathbf{R}_i)|}{|(\bigcup_i \mathbf{R}_i)|} \tag{8}$$

where low values of F_P and F_N indicate high localization accuracy.

4.3. Results on MICC-F600 dataset

As previously introduced, our method detects an image as forged if there is at least one affine transformation that fits at least between a number N of points of an image area and other N points of another image zone. The choice of the value N is crucial because it obviously impacts on TPR and FPR. In this subsection, we set up the value of N by means of ROC curves through the analysis of the performance, in terms of TPR and FPR, on the MICC-F600 dataset.

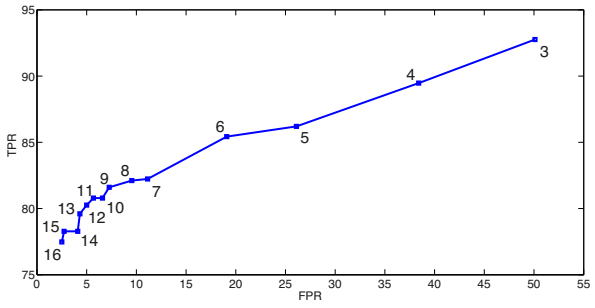

Figure 6: MICC-F600: ROC curve of TPR vs FPR varying N.

In Figure 6 we report the authenticity detection results, obtained by varying the parameter N within the interval $[3,\ldots,16]$ ($N = 3$ is the minimal number of points required for the estimate of an affine transformation). The best possible detection method would yield a point in the upper left corner, corresponding to FPR = 0% (no false positives) and TPR = 100% (no false negatives). It is to be noted that if we consider as valid, for instance, a set of points $N = 4$, the system achieves a good TPR (89.47%), but it shows a high FPR (38.4%). However, this effect is drastically reduced when we consider transformations that have a greater consensus (i.e. $N \geqslant 7$), where we achieve good performance with a still high TPR, around 80%, and a low FPR, around 6%. By using a criterion of minimum Euclidean distance from the ideal point located in the upper left corner, we can select $N = 9$ as the best value that will be used for the rest of the experiments. In fact, with this setting our algorithm gives a TPR equal to 81.6% and a FPR of 7.27%.

To support such a choice, we have also analyzed the localization performance in terms of F_P and F_N by varying the parameter N as before (see Figure 7). It can be noticed that, differently from what happened to TPR and FPR, the variability interval of F_P and F_N is quite small. In fact, F_P remains almost constant (from 0.4% to 0.24%) while F_N slightly increases (from 5.58% to 7.83%). The increment of F_N is mainly caused by the fact that with a high N some forged images are not correctly detected, so in these cases F_N is equal to 1. Based on this analysis, we can see that the previous choice to assume $N = 9$ is plausible (i.e. it returns a F_P of 0.31% and a F_N of 6.59%).

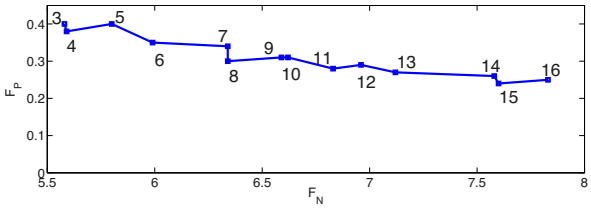

Figure 7: MICC-F600: ROC curve of F_P vs F_N varying N.

4.3.1. A new strategy for the selection of the sample set

To further augment the performance of the proposed method, we have designed, as already assessed in Section 3, an improved selection strategy to collect the minimal sample set necessary to estimate the geometric transformation between the original and the copied area. Such a new strategy is compared with that one described in Kanazawa *et al.* [28], which is used by default in the original J-Linkage algorithm. Table 1 presents the performance of our method in terms of both *authenticity detection* (TPR vs FPR) and *patch localization* (F_P vs F_N), obtained by applying the two different strategies. Results are reported for three values of the parameter N (8, 9 and 10). It is possible to point out that the new strategy basically improves the performance for all the values of N. In particular, if we consider the results obtained with $N = 9$ (the "set up value"), we can observe a growth in terms of TPR from 76.61% to 81.60%. This is due to the fact that the estimate of the geometric transformation between the original and the copied area, is more accurate and stable than in [28].

Table 1: Comparison of different sample-set sampling strategies: our method vs Kanazawa *et al.* [28].

	$N = 8$		$N = 9$		$N = 10$	
%	Ours	[28]	Ours	[28]	Ours	[28]
TPR	82.11	79.60	81.60	76.61	80.79	78.81
FPR	9.54	12.20	7.27	8.68	6.59	7.27
F_P	0.30	0.24	0.31	0.24	0.31	0.41
F_N	6.34	8.59	6.59	8.47	6.62	12.27

4.3.2. A qualitative analysis

Hereafter, we show in Figure 8 some examples to qualitatively evaluate the results achieved by the proposed method. It can be appreciated how the technique

is able to accurately segment source and destination patches, also in very challenging cases. The shapes of the copied areas are not usually regular (see for example Figure 8e, 8f and 8g). Often, the original and the cloned area are very close each other (Figure 8b, 8d and 8e). Furthermore, images contain multiple copy-pasted couples (e.g. Figure 8a and 8e) and also cases with multiple cloning (e.g. Figure 8h). Finally, it is interesting to notice that a specific situation like patch flipping (Figure 8c, in which the two horses have been flipped) is managed properly too.

4.4. Comparison with our previous method (Amerini et al. [6])

After having adequately designed and set-up the proposed method, we have performed a comparison with our previous work (Amerini *et al.* [6]), both on the basis of MICC-F600 and MICC-F2000 datasets. In fact, the two methods share the same SIFT extraction and matching procedure, but differs substantially in the clustering phase: the proposed method, based on an improved variant of J-Linkage, carries out clustering in the domain of the transformation parameters, while the second one [6] implements a classical spatial clustering in the image domain (agglomerative hierarchical clustering). Table 2, shows the results of the comparison in terms of TPR and FPR (a comparison on patch localization is not feasible because the method in [6] does not deal with that).

The proposed technique achieves superior performances: 12% of improvement for TPR and 5.2% of reduction for FPR. Such a gain is mainly due to the capacity to handle forged images containing pasted areas partially overlapped or very close to the original regions. This issue is well evidenced in Figure 9. The situation in Figure 9 (top), though clustering is not correct, does not affect detection performances because the image is rightly labeled as forged anyway; but in the case in Figure 9 (bottom), the image is wrongly assigned as authentic, being the source and destination areas too close to be distinguished as separated. It is interesting to evidence that in the case of MICC-F2000 dataset (see Table 2), the performance are high for both techniques; this confirms that such a dataset is less challenging and that the proposed method still outperforms the previous one, though with a reduced gap both in TPR and FPR.

4.5. Comparisons on patch localization with other relevant methods

To make a comparison on the issue of patch localization, we have used the SATS-130 dataset [24]. In fact,

Table 2: Comparison between our proposed method and [6].

	Dataset	Ours	Amerini *et al.* [6]
TPR (%)	MICC-F600	81.6	69.0
FPR (%)		7.27	12.5
TPR (%)	MICC-F2000	94.86	93.42
FPR (%)		9.15	11.61

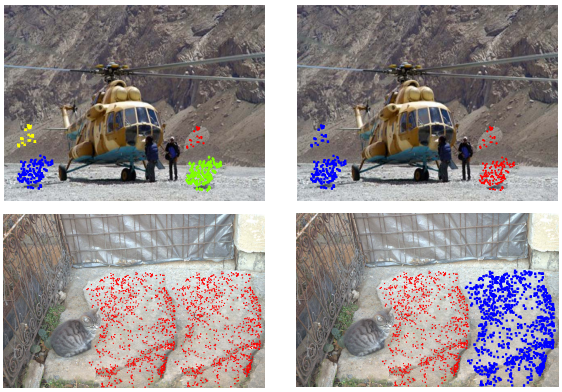

Figure 9: Two example of the results obtained with different clustering strategies: *left)* in the spatial domain (HAC), and *right)* in the transformation parameter domain (the proposed method).

such a dataset allows a complete benchmarking with several algorithms, since the authors of this dataset have reported in their paper [24] the performance in terms of patch localization (F_P and F_N) of the most relevant methods.

In Table 3, we report the results of three of these methods, with and without the usage of the SATS (Same Affine Transformation Selection) approach, as proposed by the authors in [24]. SATS is a post-processing method that can smoothly replace the widely used shift-vectors. In particular it can detect arbitrary variations in rotation and scaling in the duplicated region. These methods are claimed to be scale and rotation invariant, so they are comparable with the proposed approach; for sake of conciseness only the best performing three have been taken into account and hereafter briefly explained. The first technique (i.e. INT2 [21]) proposed a method to detect duplicated regions even when the copied portion have experienced reflection, rotation or scaling. To achieve this, overlapping blocks of pixels are resampled into log-polar coordinates, and then summed along the angle axis, to obtain a one-dimensional descriptor invariant to reflection and rotation; moreover, scaling in rectangular coordinates results in a simple translation of the descriptor. The method named INT4 [22], achieves robustness to copied region rotation by firstly reducing the image dimension through a Gaus-

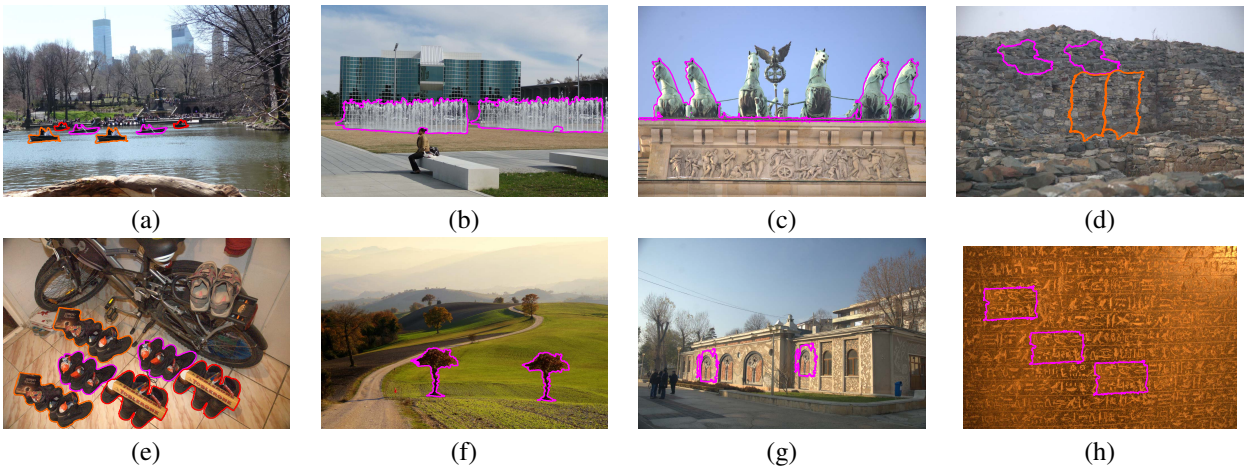

Figure 8: Examples of patch localization on the MICC-F600 dataset.

sian pyramid and then extracting four features for each circle block. The feature vectors are then lexicographically sorted and similar vectors will be matched according to a certain threshold value. Finally, the technique denominated MOM3 [23] uses the Zernike moments feature vectors to grant copy-move localization in presence of copied patch rotation.

Table 3: Test on the SATS-130 dataset: patch localization.

Method	F_P (%)	F_N (%)
INT2 [21]	4	96
INT4 [22]	24	66
MOM3 [23]	0.4	88
INT2 [21] + SATS [24]	0	22
INT4 [22] + SATS [24]	0	41
MOM3 [23] + SATS [24]	0	23
Our method	0.66	16.34

It is immediate to observe that the proposed method outperforms all the others, in both fashions (with and without SATS approach), by obtaining a F_P around zero and a F_N of 16.34%.

5. Conclusion

In this paper a new technique based on SIFT features to detect and localize copy-move forgeries has been presented. The main novelty of the work consists in introducing a clustering procedure which operates in the domain of the geometric transformation; such a procedure has been properly improved to deal with multiple cloning too. Experimental tests have been carried out on different datasets containing various typologies of fake images and also original ones. Results confirm that the proposed method outperforms other similar state-of-the-art techniques both in terms of copy-move forgery detection reliability and of precision in the localization of the manipulated patches.

Acknowledgements

This work was partially supported by the REWIND Project, funded by the FET programme under the 7FP of the EC, and by the SECURE! Project, funded by the POR CreO FESR 2007-2013 programme of the Tuscany Region (Italy).

References

[1] B. Mahdian, S. Saic, A bibliography on blind methods for identifying image forgery, Signal Processing: Image Communication 25 (6) (2010) 389–399.

[2] J. A. Redi, W. Taktak, J.-L. Dugelay, Digital image forensics: a booklet for beginners, Multimedia Tools and Applications 51 (1) (2011) 133–162.

[3] H. Farid, A survey of image forgery detection, IEEE Signal Processing Magazine 2 (26) (2009) 16–25.

[4] D. G. Lowe, Distinctive image features from scale-invariant keypoints, Int.'l Journal of Computer Vision 60 (2) (2004) 91–110.

[5] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, G. Serra, Geometric tampering estimation by means of a SIFT-based forensic analysis, in: Proc. of IEEE ICASSP, Dallas, TX, USA, 2010.

[6] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, G. Serra, A SIFT-based forensic method for copy move attack detection and transformation recovery, IEEE Transactions on Information Forensics and Security 6 (3) (2011) 1099–1110.

[7] R. Toldo, A. Fusiello, Robust multiple structures estimation with J-Linkage, in: Proc. of ECCV, Marseille, France, 2008.

[8] A. Popescu, H. Farid, Exposing digital forgeries by detecting duplicated image regions, Tech. Rep. TR2004-515, Dartmouth College, Computer Science (2004).

[9] Y. Huang, W. Lu, W. Sun, D. Long, Improved DCT-based detection of copy-move forgery in images, Forensic Science International 206 (1-3) (2011) 178–184.
[10] J. Fridrich, D. Soukal, J. Lukás, Detection of copy-move forgery in digital images, in: Proc. of DFRWS, Cleveland, OH, USA, 2003.
[11] G. Li, Q. Wu, D. Tu, S. J. Sun, A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD, in: Proc. of IEEE ICME, Beijing, China, 2007.
[12] Z. He, W. Sun, W. Lu, H. Lu, Digital image splicing detection based on approximate run length, Pattern Recognition Letters 32 (12) (2011) 1591–1597.
[13] Z. Lin, J. He, X. Tang, C.-K. Tang, Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis, Pattern Recognition 42 (11) (2009) 2492–2501.
[14] Z. He, W. Lu, W. Sun, J. Huang, Digital image splicing detection based on markov features in DCT and DWT domain, Pattern Recognition 45 (12) (2012) 4292–4299.
[15] M. Bashar, K. Noda, N. Ohnishi, K. Mori, Exploring duplicated regions in natural images, IEEE Transactions on Image Processing in press.
[16] W. Luo, J. Huang, G. Qiu, Robust detection of region-duplication forgery in digital image, in: Proc. of ICPR, Washington, DC, USA, 2006.
[17] H.-J. Lin, C.-W. Wang, Y.-T. Kao, Fast copy-move forgery detection, WSEAS Trans. Sig. Proc. 5 (5) (2009) 188–197.
[18] B. Mahdian, S. Saic, Detection of copy-move forgery using a method based on blur moment invariants, Forensic Science International 171 (2-3) (2007) 180–189.
[19] S. Bayram, H. Taha Sencar, N. Memon, An efficient and robust method for detecting copy-move forgery, in: Proc. of IEEE ICASSP, Washington, DC, USA, 2009.
[20] W. Li, N. Yu, Rotation robust detection of copy-move forgery, in: Proc. of IEEE ICIP, Hong Kong, China, 2010.
[21] S. Bravo-Solorio, A. K. Nandi, Passive method for detecting duplicated regions affected by reflection, rotation and scaling, in: Proc. of EUSIPCO, Glasgow, Scotland, 2009.
[22] J. Wang, G. Liu, H. Li, Y. Dai, Z. Wang, Detection of image region duplication forgery using model with circle block, in: Proc. of MINES, Washington, DC, USA, 2009.
[23] S.-J. Ryu, M.-J. Lee, H.-K. Lee, Detection of copy-rotate-move forgery using zernike moments, in: Proc. of Int.'l Workshop on Information Hiding, Calgary, Canada, 2010.
[24] V. Christlein, C. Riess, E. Angelopoulou, On rotation invariance in copy-move forgery detection, in: Proc. of IEEE WIFS, Seattle, WA, USA, 2010.
[25] X. Pan, S. Lyu, Region duplication detection using image feature matching, IEEE Transactions on Information Forensics and Security 5 (4) (2010) 857–867.
[26] B. L. Shivakumar, S. Baboo, Detection of region duplication forgery in digital images using SURF, International Journal of Computer Science Issues 8 (4) (2011) 199–205.
[27] P. Kakar, N. Sudha, Exposing postprocessed copy-paste forgeries through transform-invariant features, IEEE Transactions on Information Forensics and Security 7 (3) (2012) 1018–1028.
[28] Y. Kanazawa, H. Kawakami, Detection of planar regions with uncalibrated stereo using distribution of feature points, in: Proc. of BMVC, Kingston, UK, 2004.
[29] R. I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.
[30] T.-T Ng, S.-F. Chang, J. Hsu, M. Pepeljugoski, Columbia Photographic Images and Photorealistic Computer Graphics Dataset, ADVENT, Columbia University, 2004.