SIFT MATCH REMOVAL AND KEYPOINT PRESERVATION THROUGH DOMINANT ORIENTATION SHIFT

R. Caldelli[*†], *I. Amerini*[†], *A. Costanzo*[‡]

[*] Florence Research Unit
National Interuniversity
Consortium for Telecomm.
Parma, Italy

[†] Media Integration and
Communication Center
University of Florence
Florence, Italy

[‡] Siena Research Unit
National Interuniversity
Consortium for Telecomm.
Parma, Italy

ABSTRACT

In Image Forensics, very often, copy-move attack is countered by resorting at instruments based on matching local features descriptors, usually SIFT. On the other side, to overcome such techniques, smart hackers can try firstly to remove keypoints before performing image patch cloning in order to inhibit the successive matching operation. However, keypoint removal determines per se some suspicious empty areas that could indicate that a manipulation has occurred. In this paper, the goal to nullify SIFT matches while preserving keypoints is pursued. The basic idea is to succeed in altering the features descriptor by means of shifting the dominant orientation associated to a specific keypoint. In fact, to provide rotation invariance, all the values of the descriptor are computed according to such orientation. So doing, it should impair the whole matching phase.

Index Terms— SIFT, dominant orientation, copy-move attack, image forensics, edges.

1. INTRODUCTION

Scale Invariant Feature Transform (SIFT) [1] is one of the most popular Computer Vision algorithms of the last decade. Thanks to its invariance to geometrical manipulations and its robustness against several processing, the SIFT algorithm has been successfully employed in a huge variety of scientific fields, including scene (object) recognition and detection, image retrieval, image registration, image forgery detection, panorama stitching, automated navigation and tracking. In Image Forensics, the high distinctiveness of SIFT features makes them the ideal choice for matching-based forensic detection (e.g. copy-move), whereby image regions are duplicated to hide or introduce semantically relevant contents [2, 3]. Recent studies have demonstrated that it is possible for a skilled attacker to exploit the design of the SIFT algorithm (e.g. decision methods) to intentionally prevent keypoints from being detected. By doing so, the security and the correct functioning of SIFT-based applications are severely threatened. The most effective methods proposed so far to prevent the detection of SIFT matches erase one or both of the matched keypoints [4–7]. Such approaches, although quite effective, introduce forensically detectable traces by artificially depriving semantically relevant image regions of their keypoints. To tackle with this visibility issue, the dual problem of keypoint injection was studied to repopulate the attacked regions with detectable keypoints not matching with the original counterparts prior to the removal [8, 9]. Generally speaking, there could be two potential drawbacks with this strategy: firstly, following removal, the image must undergo a second attack, which may further degrade the perceptual quality; secondly, the population of injected keypoints is typically lower than that of the original keypoints. Despite the attempts to hide such manipulation, keypoint removal detectors robust to keypoint injection were devised in [10].

In this paper, we present an alternative method to delete SIFT matches that does not require the removal (and injection) of keypoints. To this aim, we propose to modify the dominant orientation of one of the matching keypoints. More specifically, we replace the neighborhood of the targeted keypoint with another neighborhood drawn from a database built in such a way that each of its elements is a patch containing one keypoint. Among all the possible substitutions, we select the new patch based on two criteria: i) the new dominant orientation must be sufficiently different from the original orientation and thus from the orientation of the second keypoint involved in the match; and ii) the two patches must be similar according to some similarity metric, so that visually unpleasant artifacts are minimized. To the best of our knowledge, the concept to perturb the dominant orientation has been studied in the context of Content-Based Image Retrieval (CBIR) [11] to delude image recognition techniques.

The outline of the paper is the following: Section 2 briefly

The authors would like to thanks Prof. Mauro Barni for his basic support in discussions on the main issues of the work.

The work has been partially supported by the SMARTVINO Project, funded by the PRAF 2012-2015-1.2.e programme of the Tuscany Region (Italy) and by the MAVEN Project, funded by the EU 7^{th}FP under grant 606058.

summarizes the principles underlying the SIFT keypoint extraction, Section 3 describes the proposed method to remove SIFT matches and Section 4 experimentally validates it; Section 5 concludes the paper.

2. SIFT AND THE DOMINANT ORIENTATION

Given an image, SIFT features [1] are detected at different scales by using a scale space representation implemented as an image pyramid, where each level is obtained by Gaussian smoothing and sub-sampling of the image resolution. Interest points, commonly referred to as *keypoints*, are selected as local extrema in the scale-space by applying the Difference of Gaussians (DoG) operator. To keep only those DoG extrema that can guarantee scale invariance, the SIFT algorithm performs two checks verifying whether the local contrast and the distance from edges (considered unstable) are sufficiently large. All the candidates that survived the above refinement process are stable keypoints. Subsequently, to each keypoint is assigned a *dominant orientation*, in such a way that the keypoint remains recognizable when the image is arbitrarily rotated. For each spatial coordinate, the gradient magnitude and orientation are computed. Then, orientations are first weighted by their magnitudes and by a circular Gaussian window and then organized into a histogram of 36 bins, each covering 10 degrees. As shown in the example of Fig. 1, the

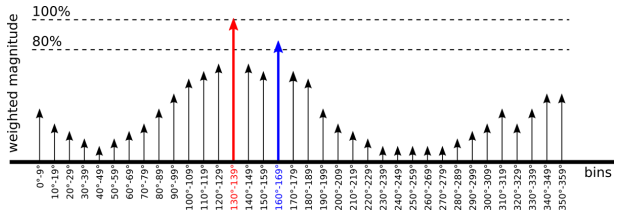


Fig. 1. Example of dominant orientation.

dominant orientation assigned to the keypoint corresponds to the highest peak of the histogram. If the histogram has other peaks whose height is comparable to that of the dominant orientation ($\geq$ 80% of the dominant peak in [1]), for each of them a new keypoint, up to a maximum of 3, is generated at the same location and scale but with different orientation. The final stage consists in computing the descriptor, that is a compact vector representation of the region surrounding the keypoint. The 16×16 patch centered on the keypoint is rotated relatively to the dominant orientation and the histograms of gradients of patch pixels are computed and concatenated into a descriptor of 128 elements.

3. THE PROPOSED METHOD

The proposed attack method substitutes a neighborhood of a keypoint, belonging to a SIFT match, with a patch containing again a keypoint but with a required dominant orientation in such a way that SIFT descriptor associated with that keypoint is altered and consequently the existing SIFT match is lost. To do that, a database of image patches ($DB_{patches}$) of squared size $w \times w$ and containing a keypoint is built up. Each patch, extracted from images taken from an as-generic-as-possible digital archive, is labeled by means of a structure $\mathbf{s_p}$ composed by three fields (see Equation (1)):

$$\mathbf{s_p} = [do, \mathbf{E}, \mathbf{P}] \tag{1}$$

where do is a double precision value which represents the dominant orientation (expressed in degrees) computed according to the SIFT procedure, $\mathbf{E}$ is a binary matrix of patch size which provides an edge description of that patch and $\mathbf{P}$ is the gray-level image patch. The attack procedure randomly selects, within the to-be-modified image, just one of the two matched keypoints (a specific selection criterion has not been implemented yet) and computes the corresponding original dominant orientation do_{orig}. According to the value of do_{orig}, the database $DB_{patches}$ is searched for all those patches whose dominant orientation do_{new} satisfies the rule in Equation (2):

$$do_{new} = do_{orig} \pm 10^{\circ} \tag{2}$$

where the sign $\pm$ is randomly chosen (for example, a criterion based on patch minimum distance could be considered). Usually a tolerance value ε is introduced ($[do_{new} - \varepsilon, do_{new} + \varepsilon]$) to avoid to obtain a limited group of candidate patches. The choice of 10° is given by the minimal requirement, as explained in Section 2, to move the dominant orientation in the nearby bin in order to shift the whole SIFT descriptor which is referenced to it. The basic idea is to induce a misalignment between SIFT descriptors involved in matching operation by reducing, as much as possible, the image distortion. To address the patch selection within the obtained (do_{new}) group of candidate patches, different methodologies can be thought; in the case at hand, two solutions have been taken into account (results will be discussed in Section 4). The first one aims at preserving the edges, in fact it tends to minimize the distance between the original patch and the candidate one according to the their binary edge description $\mathbf{E}$, while the second privileges the perceptual quality and chooses the patch with the lower distortion from the original with respect to a specific quality metric (e.g. PSNR). Finally, the selected patch $\mathbf{P}$ is substituted to the original one; this happens for all the keypoints (just one for each couple) belonging to a SIFT match. To better preserve visual quality a masking operation can be performed on the selected patch, before pasting, by combining it with the original one (original is privileged on the patch borders) according to Equation 3:

$$patch_{masked} = R * patch_{orig} + (1 - R) * patch_{selected} \tag{3}$$

where R is an empirical weighting matrix, whose elements are set to 1 along the patch borders and progressively decrease

to 0 near the center. The use or not of such an approach will be indicated in the sequel of the manuscript with the terms *mask* and *no mask* respectively.

4. EXPERIMENTAL RESULTS

In this section, experimental results to prove the effectiveness of the proposed methodology are provided; in particular, in subsection 4.1 tests on image pairs (an image and its replica) are presented, while in subsection 4.2, the case of a real forensic attack in which an image contains copy-moved patches is taken into consideration. The database $DB_{patches}$ has been created by taking 750 images from the INRIA archive[1]; for each image, keypoints have been extracted and their dominant orientations computed by using *VLFeat*[2] library (only the first octave is used). Around each keypoint, a patch of size $w \times w$ has been cut out and stored within the structure of database according to Equation (1); the edge description of the patch $\mathbf{E}$ has been obtained by applying the *Canny operator* to the patch $\mathbf{P}$. It is important to highlight that only patches containing a single keypoint are assumed as valid; two size values $w = 8, 16$ have been considered and the number of patches obtained in the two cases has been $1,094,490$ and $555,918$ respectively. Furthermore, the parameter ε has been fixed at $0.2°$ (among all the values we have checked, this granted a good trade-off between performances and complexity).

4.1. Tests on cloned images

In this set of experiments, 130 images have been randomly selected from the UCID[3] digital archive (size 384×512 pixels); each image is matched with its replica and homologue keypoints are connected in the SIFT domain. On average, such images contain less than 100 SIFT matches. In Table 1, the various obtained results are reported; two attack scenarios for the patch size ($w = 8$ and $w = 16$) are presented in the left and right side of Table 1 respectively. For each one, two choice criteria have been investigated to select the proper patch within the group of those with the required dominant orientation: the first, indicated with *PSNR*, is based on the maximization of PSNR between the original patch and the candidate one; the second, indicated with *EDGE*, minimizes the distance between the two in terms of edge description. The usage or not of the masking is clearly labeled as *mask/no-mask*. In general, it can be noted that the percentage of keypoints left in the image (keypoints belonging to a SIFT match) is averagely above 80% (first row) while removed matches (second row) are around 90%, except for the case *PSNR* and $w = 8$ in which values are around 65%. It is worthy to explain that, during the procedure, sometimes spatially close keypoints are however deleted from the successive insertion of adjoining patches. Full-frame visual quality is globally quite high: PSNR (third row) is about 56dB for the case $w = 8$ and about 50dB for the case $w = 16$, as expected. In the last row of Table 1, PSNR values computed only on patches are provided: an average value of 35dB is globally achieved. Furthermore, it can be pointed out that *EDGE* approach seems to grant a superior stability with respect to *PSNR* when changing w. To verify the performances of the proposed method, some plots for the case with $w = 16$ and *EDGE*-based selection criterion are presented in Figure 2. In particular, in Figure 2(a), the trends of the number of original matched keypoints for each image and of the left keypoints after the attack (*mask/no-mask*) are presented; on the other side, in Figure 2(b), the SIFT matches initially present with respect to the remained matches after the attack with *mask* and *no-mask* are pictured. It can be noticed that the number of keypoints is basically preserved being the three graphs almost overlapped though that representing the original keypoints a bit higher as expected. On the contrary, the remained SIFT matches are drastically reduced being the case *no-mask* lower than the *mask* one. If we make a comparison with the method presented in [11], though oriented to impair image retrieval, it can be understood that just an average local PSNR of $23.84dB$ can be obtained while a value around $35dB$ is achieved by the proposed technique.

4.2. Tests on cloned patches in a single image

In this subsection experimental results obtained by considering the real image forensic scenario where a portion of an image is copy-moved to create a fake content is debated. Experiments have been carried out to verify if the proposed methodology is able to deceive SIFT-based techniques usually adopted to reveal cloning operations. In the test reported in Table 2, 10 copy-moved attacked images ($I1 : I10$) have been taken into account. Such fake images, whose size ranges between 300×500 and 1000×1500 pixels, presents realistic forgeries obtained by duplicating one or more regions of variable sizes and shapes. In this context, results are reported for the attack configuration with parameters set at *mask* and *EDGE*. From left to right, the first column represents the number of SIFT matches detected when the simple copy-move modification is applied (obviously the number of keypoints will be double), while the following columns represent the left keypoints, the remained matches and the PSNR computed only on the modified patches for the cases $w = 8$ and $w = 16$ respectively. It can be observed that the number of remained SIFT matches, in many cases, is equal to zero, in particular when $w = 16$; averagely a percentage of only 5.52% is not deleted by the proposed method. Left keypoints and visual quality are basically the same for both cases $w = 8$ and $w = 16$, as pointed out by the average percentages (last

[1] INRIA DB http://lear.inrialpes.fr/people/jegou/data.php#holidays
[2] http://www.vlfeat.org/
[3] UCID DB: http://homepages.lboro.ac.uk/ cogs/datasets/ucid/ucid.html

	w = 8				w = 16			
	PSNR		EDGE		PSNR		EDGE	
	no-mask	mask	no-mask	mask	no-mask	mask	no-mask	mask
Left keypoints (%)	82.00	88.50	84.06	81.26	82.64	80.11	82.64	80.11
Removed matches (%)	67.65	62.08	92.94	81.56	95.90	90.23	96.79	90.68
Full-frame PSNR (dB)	56.12	59.64	51.23	56.84	49.25	52.13	46.18	51.53
Patch PSNR (dB)	35.61	39.97	31.01	37.27	33.21	36.76	30.18	36.17

Table 1. Tests on 130 UCID images: left keypoints, removed matches and visual quality.

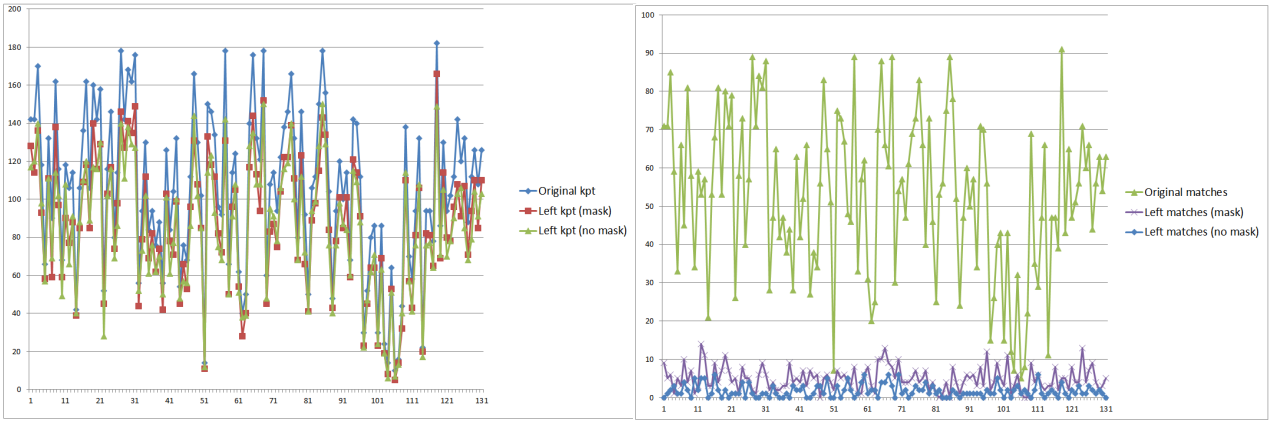


Fig. 2. Tests on 130 UCID images: left keypoints (left) and remained matches (right).

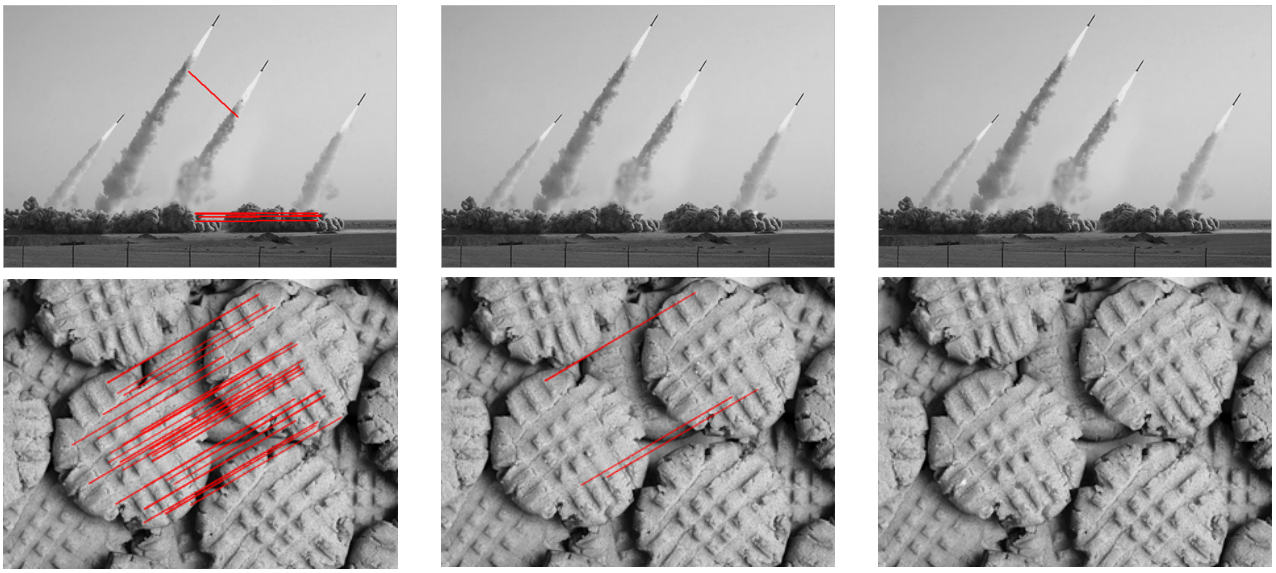
	Plain copy-move	w = 8			w = 16		
	Matches (Kpts)	Left kpt	Left matches	Patch PSNR	Left kpt	Left matches	Patch PSNR
I1	4 (8)	5	0	39.57dB	5	0	37.36dB
I2	7 (14)	13	3	39.85dB	14	0	36.89dB
I3	17 (34)	29	3	38.78dB	25	0	36.80dB
I4	14 (28)	19	2	36.86dB	26	3	37.33dB
I5	5 (10)	9	3	37.99dB	9	0	36.45dB
I6	32 (64)	60	9	38.16dB	57	6	36.11dB
I7	4 (8)	5	0	36.27dB	7	0	36.72dB
I8	124 (248)	195	17	37.73dB	209	6	37.52dB
I9	34 (68)	52	4	37.30dB	44	0	36.29dB
I10	49 (98)	84	9	37.31dB	79	1	36.13dB
AVERAGE		**81.20%**	**17.24%**	**37.98dB**	**81.90%**	**5.52%**	**36.76dB**

Table 2. Tests on copy-move attacked images: left keypoints, remained matches and local visual quality.

row of Table 2) that, anyway, are quite satisfactory. In Figure 3, to better understand the effectiveness of the proposed method, two examples (named *missiles* and *biscuits*) taken from the previous group of copy-moved attacked images, are presented. The first column contains the initial SIFT matches when a simple copy-move attack is performed, while in the other two, the cases of the proposed attack when $w = 8$ and $w = 16$ respectively are pictured. It can be pointed out that in all the circumstances, SIFT matches are completely removed except for the image *biscuits* when the used patch size ($w = 8$) is not sufficient to delete all the similarities detected between the cloned areas. When the patch size is increased to $w = 16$, no matches persist anymore.

5. CONCLUSIONS

In this paper an innovative forensic methodology which tries to inhibit SIFT matching by operating in the SIFT domain



Initial SIFT matches | Attacked 8×8 | Attacked 16×16

Fig. 3. Tests on copy-moved images *missiles* (I8) and *biscuits* (I9) (zoomed interested part). Left column represents initial SIFT matches; in central and right columns left matches after the proposed attack with $w = 8, w = 16$ are pictured respectively.

without resorting at keypoint removal-injection has been presented. This technique aims at shifting the dominant orientation which is the basic reference for SIFT description. Experimental results witnesses that a satisfactory match removal rate is achieved together with a good percentage of left keypoints and a state-of-the-art visual quality also at local level. Future works will be mainly dedicated to develop a structured and, possibly, iterative procedure, and to consider color images.

REFERENCES

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

[2] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A sift-based forensic method for copy move attack detection and transformation recovery," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 1099 –1110, sept. 2011.

[3] X. Pan and S. Lyu, "Region duplication detection using image feature matching," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 857–867, 2010.

[4] C. Y. Hsu, C. S. Lu, and S.C. Pei, "Secure and robust SIFT," in *Proc. of the 17th ACM Int. Conference on Multimedia*.

[5] T. T. Do, E. Kijak, T. Furon, and L. Amsaleg, "Understanding the security and robustness of SIFT," in *Proc. of the 18th ACM Int. Conference on Multimedia*, New York, NY, USA, 2010, pp. 1195–1198, ACM.

[6] T. T. Do, E. Kijak, T. Furon, and L. Amsaleg, "Deluding image recognition in SIFT-based CBIR systems," in *Proc. of the 2nd ACM Workshop on Multimedia in Forensics, Security and Intelligence*, New York, NY, USA, 2010, pp. 7–12, ACM.

[7] I. Amerini, M. Barni, R. Caldelli, and A. Costanzo, "Counter-forensics of sift-based copy-move detection by means of keypoint classification," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 1–17, 2013.

[8] C. S. Lu and C. Y. Hsu, "Constraint-optimized keypoint inhibition/insertion attack: security threat to scale-space image feature extraction," in *Proc. of the 20th ACM Int. Conf. on Multimedia*, 2012, pp. 629–638.

[9] I. Amerini, M. Barni, R. Caldelli, and A. Costanzo, "SIFT keypoint removal and injection for countering matching-based Image Forensics," in *The 1st ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec)*. ACM, 2013.

[10] A. Costanzo, I. Amerini, R. Caldelli, and M. Barni, "Forensic analysis of sift keypoint removal and injection," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 9, pp. 1450–1464, Sept 2014.

[11] T. T. Do, E. Kijak, L. Amsaleg, and T. Furon, "Enlarging hacker's toolbox: deluding image recognition by attacking keypoint orientations," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2012, pp. 1817–1820.