Proceedings of the Joint Austrian Computer Vision and Robotics Workshop 2020 DOI: 10.3217/978-3-85125-752-6-37

Evaluating Counter Measures against SIFT Keypoint Forensics

Muhammad Salman, Andreas Uhl
Department of Computer Sciences, University of Salzburg
uhl@cs.sbg.ac.at

Abstract. *Forensic analysis is used to detect image forgeries e.g. the copy move forgery and the object removal forgery. Counter forensic techniques (methods to fool the forensic analyst by concealing traces of manipulation) have become popular in the game of cat and mouse between the analyst and the attacker. Methods to counter forensic techniques based on SIFT keypoints are being analysed in this paper (aka anti-forensic techniques), with particular emphasis on keypoint removal in the context of copy move forgery detection. Local smoothing is suggested in this paper and turns out to be a highly attractive alternative to techniques investigated in literature so far.*

1. Introduction

In the past, images were considered as an authentic source of information – with increasing popularity and the availability of low-cost image editing software such as Adobe photoshop, corel paint shop and GIMP the truthfulness of an image can no longer be taken for granted. Among other forgery types, copy move forgery and object removal forgery are the most prominent ones. In a *copy move forgery*, a part of the image itself is copied and pasted into another part of the same image to conceal an important object or information, or to conceal that an object has been removed from the image in an *object removal forgery*. In most cases of image forgery, it is extremely difficult to distinguish between an original image and the forged one. Therefore, it is required to develop methods/techniques to assess the authenticity of an image – Digital Image Forensics (DIF [19]) has served this purpose to a large extent. Whenever an image is forged, there are some traces which are left behind in the forged image. These traces are useful for the forensic researcher to detect a forgery.

A wide range of DIF forgery detection techniques have been established in the recent years [4, 6, 21]. Besides recent deep learning based schemes, techniques relying on Scale Invariance Feature Transform (SIFT) keypoints have been shown to be effective. In particular, SIFT keypoints [12] have been proposed to reveal copy move forgeries [6] and image cloning [17], as well as to detect copyrighted material using CBIR techniques [9].

Attackers are making it difficult to apply these techniques by developing counter forensic techniques, i.e. by minimising those traces left behind in forged images. In the context of SIFT keypoint forensics, this is done by manipulating SIFT keypoints, e.g. removing existing ones or injecting fake key points to fool the forensic techniques. This paper is a contribution to such counter forensic approaches against SIFT-keypoint forensic techniques. In particular, we focus on SIFT keypoint removal techniques. Section 2 reviews corresponding techniques as proposed in literature and suggest a new approach. Section 3 is devoted to an extensive empirical evaluation, looking at the tradeoff among image quality, keypoint removal effectiveness as well as the generation of new keypoints. In the conclusion we discuss results obtained and give an outlook to further work in this direction.

2. SIFT Keypoint Removal Techniques

The simplest approach, *global smoothing* (GS), reduces the potential keypoints at the level of difference of Gaussian (DoG) by Gaussian smoothing (which flattens the pixel values of an image), e.g. [1] applies a Gaussian filter with $\sigma = 0.7$ and window size 3×3 as a good compromise between amount of deleted keypoints and overall visual quality of an image. A more sophisticated approach is to first apply GS (the original paper [9] suggests to employ $\sigma = 1.3$), detect remaining keypoints, and apply *local smoothing* (LS) in patches around detected key-

points, with size 3×3 to 7×7 pixels (denoted as GS+LS).

Another strategy to remove SIFT keypoints is the *collage attack* (CA) [10], which substitutes an original image patch (patch containing a keypoint) with another patch (containing no keypoint) of the same size contained in a pre-computed patch dictionary. The new patch must not contain SIFT keypoints and should be as similar as possible to the original one according to some similarity criteria (e.g., [1] created a dictionary of about 120,000 patches and chose histogram intersection distance, widely used in image retrieval applications [22], as a patch similarity measure. The same approach is used in experiments of [3].

Removal with minimum distortion (RMD) [9] adaptively calculates a small image patch and adds it to the neighbourhood of the key point such that the overall operation results in a minimum least-square distortion in the keypoint neighbourhood under the condition that the keypoint is removed. Finally, the *classification based attack* (CLBA) presented by [1] arranges GS+LS, CA, and RMD into an iterative procedure which first detects SIFT keypoints, classifies them into distinct classes, and subsequently applies one of the three individual removal techniques to the suited classes.

For all these techniques, [2] suggested to remove only one of the matching keypoints from each matching keypoints pair in case of preventing to detect copy move forgeries. There are also forensic techniques to counter those anti-forensic keypoint removal methods (see e.g. [7, 16].

As GS has significant impact on image quality (as we shall see as well in the next section), also the combination with LS (i.e. GS+LS) is affected by this quality impact. Therefore, we introduce a new technique to remove SIFT keypoints called *local smoothing* (LS), and compare the various performance indicators to already existing (smoothing) techniques i.e. GS, GS+LS, and CA.

3. Experiments

3.1. Experimental Settings

With respect to software and tools, we mainly used Matlab 2014a [14] (on Windows 7 64bit) with some internal toolboxes (parallel toolbox for fast computation, image processing toolbox) and the external library *vl_feat* [20] (the latter to smooth images and to compute SIFT keypoints; we have chosen *Edge Thresh = 12* to control the number of keypoints used). For the computation of image quality metrics (IQM) (*PSNR, VSNR, UQI, SSIM*), we used the MaTrix MuX visual quality assessment package [15]. As experimental data, we used the first 100 images (i.e. from *ucid00001.tif* to *ucid000100.tif*) from the Uncompressed Image Database (UCID) [18] for experiments for keypoint removal methods assessment. For the CA, we created a keypoint-free patch dictionary from all images using overlapping patches.

For experiments with respect to detecting actual copy move attacks, we combined two datasets to result in 100 images (50 actually forged images and 50 original images). Forged images are taken from a public dataset for assessing forensic techniques [5] (see Fig. 5 for examples), which contain simple translated copies of objects/regions, while the "original" images are taken from the RAISE dataset [8] from the `BUILDING PHOTO` category (see Fig. 6). The latter data has been included to determine the methods' robustness against indicating false positives[1]. In keypoint removal for countering copy move detection, we removed only one keypoint from each matching pair of keypoints as suggested.

3.2. Experimental Results

In order to assess the quality of the image after removing keypoints, we used different IQM, i.e. `PSNR, SSIM, VSNR, and UQI`. Fig. 1 compares three different techniques, i.e. GS, GS+LS, and LS. In GS+LS, an image is smoothed first globally with $\sigma = 1.3$ as suggested in literature and afterwards patches containing keypoints (of different sizes) are smoothed locally. In the plots, different smoothing strength (different σ values) is depicted on the **X axis**, while the **Y axis** represents the output value for a specific image quality measure.

Fig. 1 reveals that the quality of a locally smoothed (`LS`) image is better in comparison to the other two smoothing techniques (i.e. `GS` and `GS+LS`) for all IQM. GS deteriorates image quality quickly for increasing smoothing strength. Also for the combined method GS+LS the quality is found to be rather low due to the impact of GS. The quality of the LS images is better because we are smoothing only the patches around SIFT keypoints while other pixels are left untouched. As expected, when increasing the patch size in LS and GS+LS, the quality of the pro-

[1] Similar looking structures within an image may lead to an image incorrectly being classified as copy move forged image.

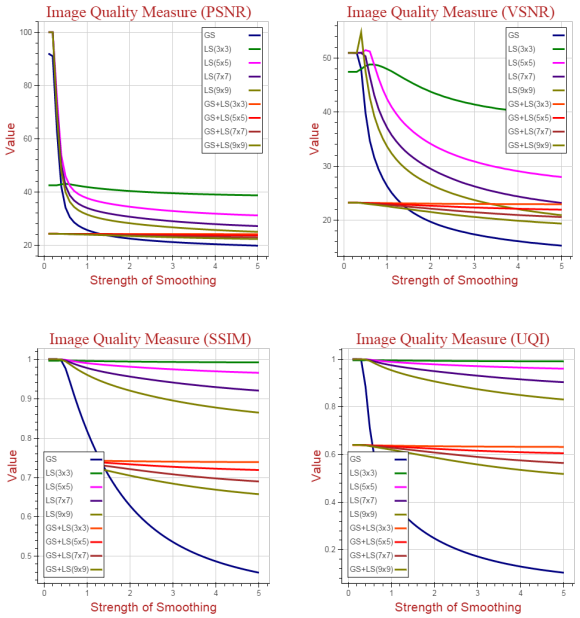


Figure 1: IQM Comparison among GS vs LS vs GS+LS

cessed images decreases.

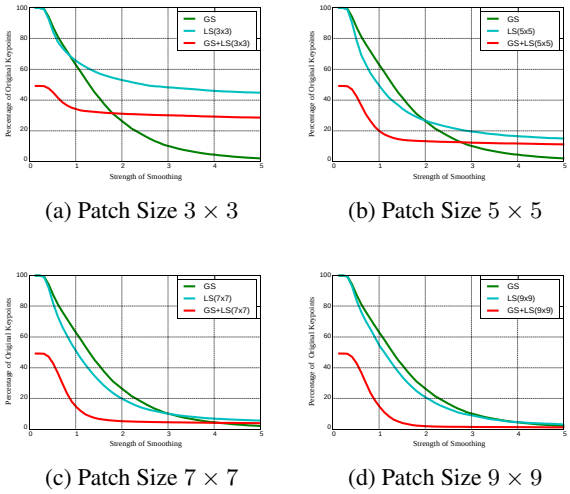
Table 1 displays IQM values for the CA.

Patch Sizes	PSNR	VSNR	UQI	SSIM
3x3 Patch	64.80	32.78	0.99	0.99
5x5 Patch	51.95	32.23	0.99	0.99
7x7 Patch	47.10	47.58	0.99	0.99
9x9 Patch	42.86	40.89	0.98	0.99

Table 1: IQM for CA.

For UQI as well as SSIM we notice almost no quality degradation by the CA, no matter which patch size is being used. For PSNR, CA is superior to all GS+LS variants and for almost all other settings except for extremely low smoothing strength. Finally, for VSNR, CA is again superior to all GS+LS variants and for all other techniques but LS with patchsize 3 for low smoothing strength. Overall, the quality obtained with the CA is very good, and only comparable to LS with patchsize 3, however, with all patch sizes considered.

But how effective are the smoothing-based methods in actually removing keypoints ? Contrasting to CA, in which all present keypoints are replaced by keypoint-free patches, smoothing does not guarantee that keypoints are actually removed. Fig. 2 illustrates the percentage of original keypoints which are still present after smoothing for increasing smoothing strength.



(a) Patch Size 3×3 (b) Patch Size 5×5

(c) Patch Size 7×7 (d) Patch Size 9×9

Figure 2: Share of retained keypoints: GS vs LS vs GS+LS

For larger patch sizes, GS and LS perform almost identically (which is clear considering the definition), while GS+LS is most effective in removing keypoints. For smaller patch sizes, GS is most effective for high smoothing strength, while GS+LS is best for low smoothing strength. LS is not very effective under these conditions.

When applying techniques for keypoint removal, new keypoints are being created, e.g. at the edge of the patches in CA, LS, and GS+LS. This is not desired, as these new keypoints might match to existing ones and thus aid the forensic analyst. Fig. 3 illustrates the creation of new, additional keypoints by showing the percentage of newly created ones. LS clearly introduces the lowest number of additional keypoints, and if the size of the smoothing patch is increased then also the number of new keypoints is also increased. The smoothing strength also plays a certain role: For weak smoothing, increasing the strength leads to more new keypoints, while after reaching a peak, a further increase of smoothing strength decreases the number of newly created keypoints. This effect is expected and most obvious for GS.

In Table 2, the percentage of newly created keypoints for CA is shown. Only LS with patchsize 3 gives better results, for all other techniques we notice higher percentages of newly created keypoints when comparing Fig. 3 to the values in Table 2.

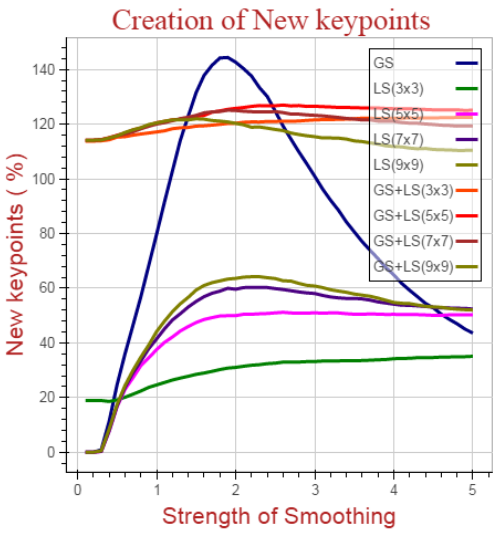


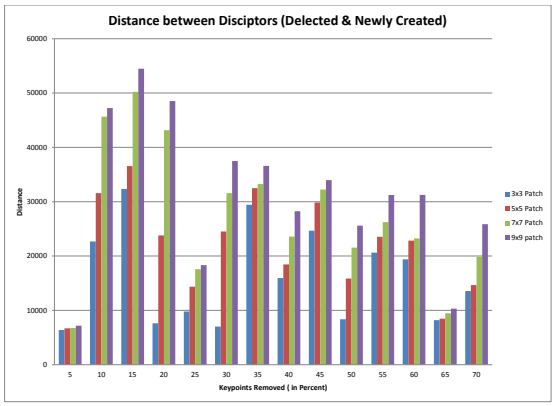
Figure 3: Creation of New Keypoints.

3x3 Patch	5x5 Patch	7x7 Patch	9x9 Patch
43.01%	39.76%	34.43%	32.21%

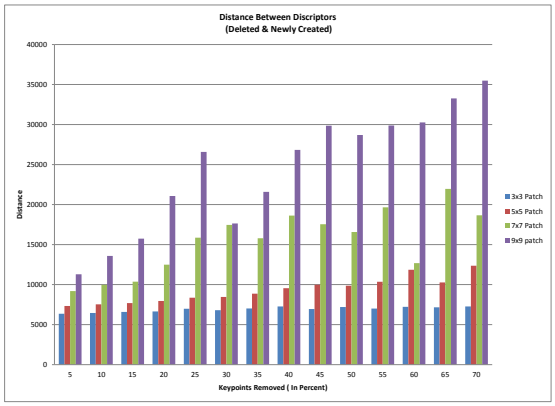
Table 2: Newly Generated Keypoints in CA.

When new keypoints are being generated, it is not their number that is most important. The aim of removing keypoints is compromised, if the newly generated ones are similar to the removed ones in terms of their SIFT descriptors. In this case, attacks might still be recovered by the forensic analyst even though keypoints have been removed. In Fig. 4 we plot the distance (squared Euclidean distance (SED)) of the SIFT descriptors describing removed and newly created ones. In particular, we compute SED between removed keypoints and their closest newly generated keypoints in terms of their descriptors. To avoid bias, we divide the result by the number of removed keypoints, as we display results in terms of increasing percentage of removed keypoints.

For the patch-based techniques, an increase of the patch size leads to higher SED, which is expected and desired. When increasing the percentage of removed keypoints, there is a tendency for increasing SED, except for LS and CA with smaller patch sizes. The largest SED values (which is the aim when removing keypoints) are seen for techniques involving GS (not shown) when a large share of all keypoints is being removed. CA clearly exhibits the lowest values, which means that the advantage of this approach in removing all keypoints is endangered by the creation of new keypoints which are close to the



(a) LS



(b) CA

Figure 4: Distance to newly created keypoints.

removed ones in terms of their SIFT descriptors.

After having analysed four different SIFT keypoint removal techniques with respect to different properties, we tested these methods in an actual copy move forgery scenario. The following definitions are employed:

- $TruePositive(TP)$: A true positive test result for a forged image is one that detects at least τ matching keypoint pairs.
- $FalseNegative(FN)$: A false negative test result for a forged image is one that detects at most $\tau - 1$ matching keypoint pairs.
- $TrueNegative(TN)$: A true negative test result for an image from the `BUILDING PHOTO` category is one that detects at most $\tau - 1$ matching keypoint pairs.

- $FalsePositive(FP)$: A false positive test result for an image from the BUILDING PHOTO category is one that detects at least τ matching keypoint pairs.

Based on these definitions, we are able to compute *precision*, *recall*, and *F1-score*. Recall, that the aim of the attacker is to disable the techniques of the forensic analyst. Thus, the attacker developing these techniques to counter SIFT keypoint based forensic techniques by removing keypoints aims for low TP (and low TN), as high FN makes the forensic analyst miss forged images and high FP confuses the analyst as many genuine images are determined as forgeries.

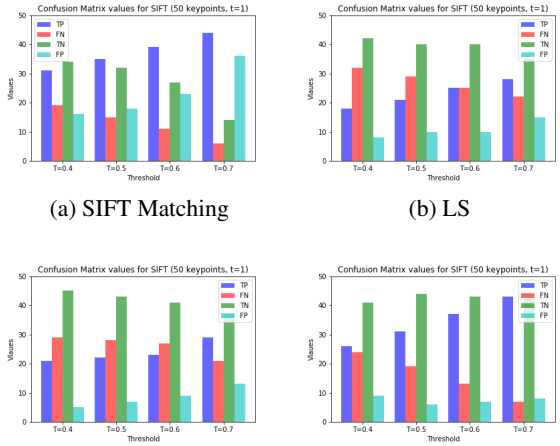
First, we computed SIFT keypoints and then for each keypoint we found the two nearest neighbours from all remaining keypoints using a K-d tree based on Euclidean distance d_1 and d_2 (where d_1 and d_2 are distances and d_1 corresponds to the closest neighbour), $T \in (0, 1)$. [13] and [11] suggested that there is a match only if $\frac{d_1}{d_2} < T$ holds. In these papers $T = 0.6$ but we looked into results for $T = 0.4$, $T = 0.5$, $T = 0.6$ and $T = 0.7$.

Figure 5: Forged Images

Figure 6: Original Images

Fig. 7 shows confusion matrices (i.e. the number of TP, FN, TN, FP) for using 50 keypoints, $\tau = 1$, for four different values of T, comparing copy move forgery detection without manipulating images, and with applying keypoint removal techniques LS, CA, and GS+LS. Patch size is set to 9x9 pixels in all patch-based techniques.

Overall, we observe that all three SIFT keypoint removal strategies work, i.e. they reduce significantly the number of TP. However, they increase also the number of TN, thus, the number of false positives is also reduced (which is not desired). When we



(a) SIFT Matching (b) LS

(c) CA (d) GS+LS

Figure 7: Copy Move Forgery Detection

compare the three removal strategies, GS+LS clearly has a higher number of TP, thus is least efficient and does not need to be considered further in this comparison. LS and CA are close, with slight advantages for LS, however, difficult to confirm in this visual representation.

When looking into recall and precision values for $\tau = 1, 2, 3$ and $T = 0.4, 0.5, 0.6, 0.7$ using 50, 100, and 200 keypoints (overall 36 configurations), we find precision(LS) $<$ precision(CA) in 33/36 cases, while recall(LS) $<$ recall(CA) in 20/36 cases. Therefore, overall, LS is clearly more effective in preventing to detect a copy move forgery as CA is. In terms of F1-score F1(LS) $\leq$ F1(CA) in 27/36 cases, which confirms the trend.

Table 3 shows precision, recall and F1-scores of the confusion matrices shown in Fig. 7. The cases in which LS delivers the best (lowest) results are underlined - we notice that this is also the clear majority within these result subsets.

		CA			LS			GS+LS		
τ	T	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
1	0.4	0.81	0.42	0.55	0.69	0.36	0.47	0.86	0.60	0.71
1	0.5	0.76	0.44	0.56	0.68	0.42	0.52	0.81	0.62	0.70
1	0.6	0.72	0.46	0.56	0.71	0.50	0.57	0.84	0.74	0.79
1	0.7	0.69	0.58	0.63	0.65	0.56	0.60	0.84	0.86	0.85

Table 3: Comparison of keypoint removal techniques in terms of precision, recall, and F1-score.

4. Conclusion

Local smoothing (LS), as proposed in this paper, turns out to be more effective in preventing a detection of a copy move attack as compared to the col-

lage attack (CA). For the patch-size chosen in the comparison, the image quality is slightly superior for CA. GS and GS+LS as also proposed in literature are neither competitive in terms of maintained image quality nor in terms of preventing the copy move attack detection capability. When considering the ease of application, LS is clearly preferable, as CA requires the generation of a keypoint-free dictionary and a vector-quantisation like patch selection process, while LS only applies a local Gaussian smoothing. Overall, LS turns out to be a highly attractive alternative to SIFT keypoint removal techniques applied so far in literature.

References

[1] 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*, 2013(1):18, 2013.

[2] I. Amerini, M. Barni, R. Caldelli, and A. Costanzo. Removal and injection of keypoints for sift-based copy-move counter-forensics. *EURASIP Journal on Information Security*, 2013(1):8, 2013.

[3] I. Amerini, F. Battisti, R. Caldelli, M. Carli, and A. Costanzo. Exploiting perceptual quality issues in countering sift-based forensic methods. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2664–2668. IEEE, 2014.

[4] E. Ardizzone, A. Bruno, and G. Mazzola. Detecting multiple copies in tampered images. In *2010 IEEE International Conference on Image Processing*, pages 2117–2120. IEEE, 2010.

[5] E. Ardizzone, A. Bruno, and G. Mazzola. Copy–move forgery detection by matching triangles of keypoints. *IEEE Transactions on Information Forensics and Security*, 10(10):2084–2094, 2015.

[6] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou. An evaluation of popular copy-move forgery detection approaches. *IEEE Transactions on Information Forensics and Security*, 7(6):1841–1854, 2012.

[7] A. Costanzo, I. Amerini, R. Caldelli, and M. Barni. Forensic analysis of sift keypoint removal and injection. *IEEE Transactions on Information Forensics and Security*, 9(9):1450–1464, 2014.

[8] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato. Raise: a raw images dataset for digital image forensics. In *Proceedings of the 6th ACM Multimedia Systems Conference*, pages 219–224. ACM, 2015.

[9] T.-T. Do, E. Kijak, T. Furon, and L. Amsaleg. Deluding image recognition in sift-based cbir systems. In *Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence*, pages 7–12. ACM, 2010.

[10] C.-Y. Hsu, C.-S. Lu, and S.-C. Pei. Secure and robust sift. In *Proceedings of the 17th ACM International Conference on Multimedia*, pages 637–640. ACM, 2009.

[11] H. Huang, W. Guo, and Y. Zhang. Detection of copy-move forgery in digital images using sift algorithm. In *Pacific-Asia Workshop on Computational Intelligence and Industrial Application, 2008. PACIIA'08.*, volume 2, pages 272–276. IEEE, 2008.

[12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[13] B. Mahdian and S. Saic. Detection of copy–move forgery using a method based on blur moment invariants. *Forensic Science International*, 171(2-3):180–189, 2007.

[14] MATLAB. *version 8.3.0.532 (R2014a)*. The MathWorks Inc., Natick, Massachusetts, 2014.

[15] M. MuX. *MeTriX MuX version 1.1*. 2014.

[16] J. Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3):61–74, 1999.

[17] M. Saleem. A key-point based robust algorithm for detecting cloning forgery. In *IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, volume 4, pages 2775–2779, 2014.

[18] G. Schäfer and M. Stich. UCID - an uncompressed colour image database. *Proc. SPICE. Storage and Retrieval Methods and Applications for Multimedia*, 11(1):472–480, 2004.

[19] H. Sencar and N. M. (Eds.). *Digital Image Forensics: There is more to a picture than meets the eye*. Springer Verlag, 2012.

[20] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, pages 1469–1472, New York, NY, USA, 2010. ACM.

[21] M. Zandi, A. Mahmoudi-Aznaveh, and A. Mansouri. Adaptive matching for copy-move forgery detection. In *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 119–124. IEEE, 2014.

[22] D. Zhang and G. Lu. Evaluation of similarity measurement for image retrieval. In *International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003*, volume 2, pages 928–931. IEEE, 2003.