IMAGE FORGERY DETECTION THROUGH RESIDUAL-BASED LOCAL DESCRIPTORS AND BLOCK-MATCHING

Davide Cozzolino, Diego Gragnaniello, Luisa Verdoliva

DIETI, Universitá Federico II di Napoli, Naples - Italy

ABSTRACT

We propose a new image forgery detection technique which fuses the outputs of two very diverse tools, based on machine learning and block-matching, respectively. The machine-learning tool builds upon some local descriptors recently proposed in the steganalysis field, which are selected and merged based on an *ad hoc* measure of reliability. The block-matching tool leverages on the patchmatch algorithm for fast search of candidate matchings. Both tools are fine-tuned so as to optimize their fusion which, in turn, exploits the respective strengths and weaknesses of each tool. The proposed technique ranked first in phase 1 of the first Image Forensics Challenge organized in 2013 by the IEEE Signal Processing Society.

Index Terms— Digital forensics, forgery detection, machine learning.

1. INTRODUCTION

Digital image forensics is gaining a great deal of attention in the scientific community, and image forgery detection is probably one of the hottest topics in this field. A large number of approaches have been proposed in recent years, and not always they are tested on public dataset or the source code is made available to guarantee reproducible research. In these conditions, it is difficult to assess objectively such methods and figure out their performance in real-world applications. Driven by these considerations, the IEEE Information Forensics and Security Technical Committee (IFS-TC) launched a detection and localization forensics challenge, the First Image Forensics Challenge, with three main goals [1]

- to provide the community with an open data set and protocol for evaluation of the latest forensics techniques
- to evaluate the current state-of-the-art techniques with respect to their ability to detect image forgeries
- to set forth a standardization protocol as a common comparison ground truth for new techniques

In this paper[1] we describe the strategy we followed to tackle phase 1 of the Challenge, devoted to image forgery *detection*.

The challenge comprises several original images captured from different digital cameras with various scenes either indoor or outdoor. No information was provided on the number and types of cameras. The forged images comprise a set of different manipulation techniques such as copy-move and splicing with different degrees of photorealism.

[1]The present paper extends the technical report presented at the IEEE International Workshop on Information Forensics and Security (WIFS) held in Guangzhou in November 2013, in the special session devoted to the Forensics Challenge.

Given the nature of the dataset, we realized very soon that a fusion of different tools was necessary. Indeed, real-world image forgery detection can be extremely challenging because of the wide availability of powerful photo-editing tools which allow for different types of manipulations, and considering the large variety of operative conditions encountered in practice, including compression, blurring, distortions, etc. No single method can be expected to work satisfactorily in all these cases, and in fact the literature confirms [2, 3] that a suitable fusion of tools can largely improve detection performance over single methods, especially in adverse and unpredictable conditions.

In recent years, we have developed image forensics techniques based on camera sensor noise, a.k.a. PRNU (photo response non-uniformity) noise [4, 5, 6], and on local descriptors [7, 8]. Therefore, we decided to follow both these approaches, with the aim of fusing decisions at the end of the process. In addition, we included a third line of development based on block matching even though this is applicable only to a fraction of the forged images, those presenting copy-move forgeries.

Unfortunately it was very soon clear that the PRNU-based approach was bound to be of little use. Lacking any information on the cameras used to take the photos, we had to cluster the images based on their noise residuals. This lowered significantly the reliability of this tool which had to be eventually discarded in the fusion. On the contrary, techniques based on local descriptors appeared from the beginning very promising, and we pursued actively this line of development, drawing also from the relevant literature in the steganalysis field. Complementing such techniques with the copy-move detector, tuned so as to guarantee very high specificity, led us eventually to obtain very good results.

In the next two Sections we will describe the proposed splicing detector based on local descriptors and copy-move detector based on fast matching. Then, Section 4 describes the decision fusion strategy and the numerical results. Eventually, we draw conclusions in Section 5.

2. SPLICING DETECTION

Several feature-based techniques have been proposed in the last decade for splicing detection. Major efforts have been devoted to find good statistical models for natural images in order to select the features that guarantee the highest discriminative power. Often, in order to capture more meaningful statistics, transform-domain features have been used, as in [9] where the image undergoes block-wise discrete cosine transform (DCT) with various block sizes and first-order (histogram based) and higher-order (transition probabilities) features are collected and merged. Given the good results obtained in terms of detection accuracy, an expanded Markov-based scheme in DCT and DWT domains is followed in [10]. Interestingly,

the method proposed in [9] was inspired by prior work carried out in steganalysis which, despite the obvious differences with respect to the forgery detection field, pursues a very similar goal, that is, detecting seemingly invisible alterations of the natural characteristics of an image. A Markov-based approach has been also recently used in [11].

The same path is followed in the forgery detection technique proposed in [12], based on an approach proposed for steganalysis in [13, 14]. The major contribution consists in deriving the features based on some co-occurrence matrices computed on the thresholded prediction-error image, also called *residual image*. In fact, modeling the residuals rather than the pixel values is very sensible in these low-level methods (not based on image semantic), since the image content does not help detecting local alterations and should be suppressed altogether. In the context of forgery detection, in particular, considering that splicing typically introduces sharp edges, it is reasonable to characterize statistically some *edge image*, which can also be the output of a simple high-pass filter, like a derivative of first order. As a further advantage, the residual image has a much narrower dynamic range than the original one, allowing for a compact and robust statistical description by means of co-occurrences.

The processing path outlined above, already proposed in [13], can be therefore summarized in the following steps

1. computation of the high-pass residuals;
2. truncation and quantization;
3. feature extraction based on co-occurrence matrices of selected neighbors;
4. design of a suitable classifier on the training set.

Given its compelling rationale, and the promising results obtained in the literature, we chose to adhere strictly to this path. Even so, a large number of design choices are necessary, beginning from the high-pass filter, to end with the classifier, which impact heavily on the performance and require a lengthy development and testing phase. Fortunately, we could rely on the precious results described in a recent work on steganalysis [15], where a large number of models have been considered, analyzed, and made available online to the research community [16]. Specifically, in [15] a number of different high-pass filters have been considered, both linear and nonlinear, with various supports, different quantization and truncation strategies for the residues have been implemented and, based on some preliminary experiments, the use of some selected groups of neighbors for co-occurrence computation has been suggested. There is no doubt, as the Authors of [15] themselves point out, that better design choices are possible and should be pursued when aiming at different goals, but the wealth of models they provide allows for the rapid development and optimization of a specific processing chain, which can be then improved, as we did in this research, under some specific respects.

2.1. Implemented method

In [15] 39 different high-pass filters are proposed, which work on the grayscale version of the original image obtained by standard conversion. All such filters are quite simple, since their goal is to highlight minor variations w.r.t. typical behaviors. Two examples among the simplest are the first order horizontal *linear* filter

$$r_{i,j} = x_{i,j+1} - x_{i,j}$$

and the first order symmetric *nonlinear* filter

$$r_{i,j} = \min[(x_{i,j+1} - x_{i,j}), (x_{i+1,j} - x_{i,j})]$$

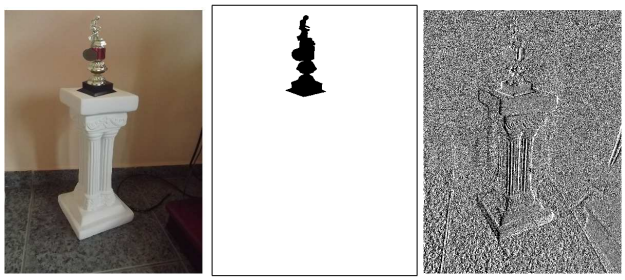

Fig. 1: A training image with its ground truth and an example residual image.

Fig.1 shows the effect of applying one of such filters to a training image of the challenge.

Residuals are in general real-valued and, although typically small, are defined on a wide range. To enable their meaningful characterization in terms of co-occurrence they must be quantized and truncated. Following [15] we use

$$\widehat{r}_{ij} = \mathrm{trunc}_T(\mathrm{round}(r_{ij}/q))$$

with q the quantization step and T the truncation value. We use $T = 2$ to limit the matrix size and consider exclusively $q = 1$, both to reduce complexity, and to limit the risk of overfitting to our training set. Each quantized residual can eventually take on 5 values, from -2 to +2. We then compute co-occurrences on four consecutive pixels along the same row or column, obtaining 625 entries, which can be highly reduced thanks to symmetries.

In the classification phase we depart significantly from the reference technique, due to the overfitting problem mentioned before. In fact, each individual model comprises 169 features for linear filters and 325 for non linear ones, a number large but still manageable with the training set available in the challenge, comprising about 1500 images (450 fake and 1050 pristine). Merging all models, however, would lead to a much larger number of features, too large to carry out a meaningful training. In [15] this problem was dealt with by means of an *ensemble* classifier, but the training set was about ten times larger.

We decided therefore to test each model individually, relying heavily on cross validation to gain a reasonable insight into their actual performance. In each experiment, we selected at random 5/6 of the pristine images and 5/6 of the fake ones to train a SVM classifier. The remaining images of each class were then used to test the trained classifier. To reduce randomness, each experiment was repeated 18 times, selecting the training and test set at random, and results were eventually averaged. Fig.2(top) shows the results for the 39 models considered, in terms of expected score, defined as

$$S = \frac{\Pr(\widehat{F}|F) + \Pr(\widehat{P}|P)}{2}$$

with $P[F]$ indicating the event "image pristine[fake]" and $\widehat{P}[\widehat{F}]$ the event "decision pristine[fake]", respectively. For several models the predicted score is in the order of 94%, hence very promising. To further improve results, we tried to merge the features of a limited number of models, up to four, not to exceed the number of training images. Results are reported in Tab.I in terms of score obtained before and after merging. The merging does not seems to guarantee any improvement over the best single-model classifier, moreover, the score exhibits a non-monotonic behavior as more models are merged, ringing an alarm bell on stability.

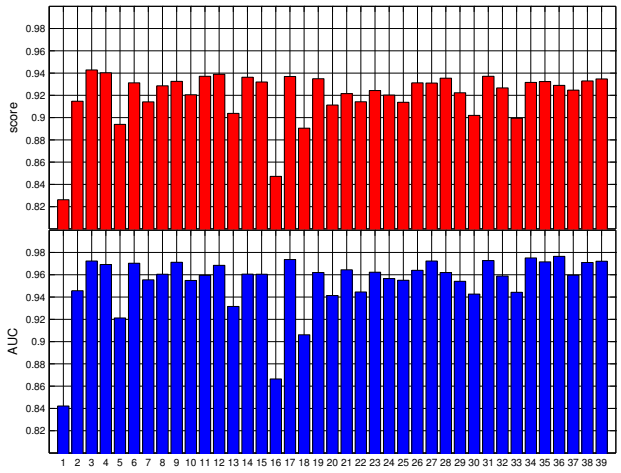

Fig. 2: Scores (top) and AUC (bottom) for all models.

To improve robustness, we considered a different measure of performance. For each SVM classifier, we displaced the separating hyperplane along the orthogonal direction, and built the corresponding ROC. Then we computed, for each model, the Area Under the receiver operating Curve (AUC), because a large AUC implies not only a good performance in the best operating point, but also robustness w.r.t. changing conditions. Fig.2(bottom) shows results. Although there is a clear correlation, the top-score models do not coincide with the top-AUC models. We then tried merging the best models selected with this latter criterion, obtaining the results reported in Tab.II. This time, performance improves monotonically with merging, providing a gain of about 1% over the best individual model,

Eventually, our SVM classifier uses the merging of all the features of models 17, 31, 34 and 36, and is trained over the whole phase-1 training set.

Model	Type	Score	AUC	Score/merg.
3	NL, 1st order	0.9429	0.9724	0.9429
4	NL, 1st order	0.9403	0.9693	0.9154
12	NL, 2nd order	0.9389	0.9685	0.9415
11	NL, 2nd order	0.9371	0.9595	0.9163

Table 1: Score obtained by the top-score individual models, and by their merging.

Model	Type	Score	AUC	Score/merg.
36	linear, 3rd order	0.9289	0.9765	0.9289
34	linear, 1st order	0.9316	0.9751	0.9462
17	NL, 3rd order	0.9369	0.9736	0.9481
31	NL, square 5$\times$ 5	0.9371	0.9727	0.9531

Table 2: Score obtained by the top-AUC individual models, and by their merging.

3. COPY-MOVE DETECTION

Many algorithms for copy-move forgery detection have been proposed in the literature [18]. They are typically based on feature matching: a search is carried out over the whole image to discover identical or very similar regions, which may be therefore due to a copy-move forgery. The basic approach is to scan all blocks of the image in sliding-window fashion [19, 20] which, however, may require a very large processing time. A popular alternative consists in finding in advance some keypoints, typically associated with major image structures, and match only the feature vectors associated with them. Leveraging on the distinctive geometrical structure of such keypoints [21], the features may be invariant w.r.t. various types of distortions, thus increasing the robustness of the matching. Nonetheless, since keypoints cannot be extracted in homogeneous regions of the image, all copy-moves involving this kind of regions, *e.g.,* hiding an aircraft by copying a fragment of sky, remain fully undetected. More in general, it has been shown experimentally [18] that techniques based on *dense* nearest neighbor fields (NNF) provide a higher accuracy, therefore we focused on this class.

A simple and pretty general detection algorithm based on this approach might comprise the following steps

1. computation of a dense NNF;
2. segmentation of the field in regions characterized by homogeneous displacement vectors;
3. selection of pairs of candidate matching regions;
4. elimination of wrong candidates based on matching error, and other criteria.

The last step is especially important as it modulates the trade off between missing detections and false alarms. If the copy-moves involves a mere translation of regions, with no further processing, any forgery of reasonable size can be detected easily, since it is very difficult to find *identical* regions in a pristine natural image. On the other hand, when some forms of processing takes place, such as resizing, rotation, change of intensity, compression, the original and copied regions might differ significantly. Therefore, by setting a low threshold on matching error we detect only a part of all possible copy-moves, those with little or no distortion but, on the other hand, eliminate false alarms. With a larger threshold, more copy-moves are detected, but the risk of false alarms increases considerably.

3.1. Implemented method

As outlined before, our first processing step is the computation of a dense NNF based on block matching. Performing the exact computation for each block of the image is exceedingly burdensome, so we resort here to PatchMatch, an iterative algorithm recently proposed for image editing applications [23, 24]. Patchmatch provides a very accurate and regular NNF, but we chose it primarily for its rapid convergence, which makes it about 100 times faster than exact methods, allowing us to process in reasonable time a large database of images.

We use 7$\times$7 pixel patches, a size that guarantees a good compromise among accuracy, resolution and speed. All image pixels are preliminarily adjusted to unitary norm, in order to single out copy-moves also in the presence of some intensity adjustments. After computing the NNF, we carry out a filtering on both horizontal and vertical components of the NNF to identify regions with homogeneous displacement. Choosing an appropriate prediction filter, we can also identify regions where displacement vectors slowly increase or decrease linearly, thus identifying also copy-moves with moderate resizing.

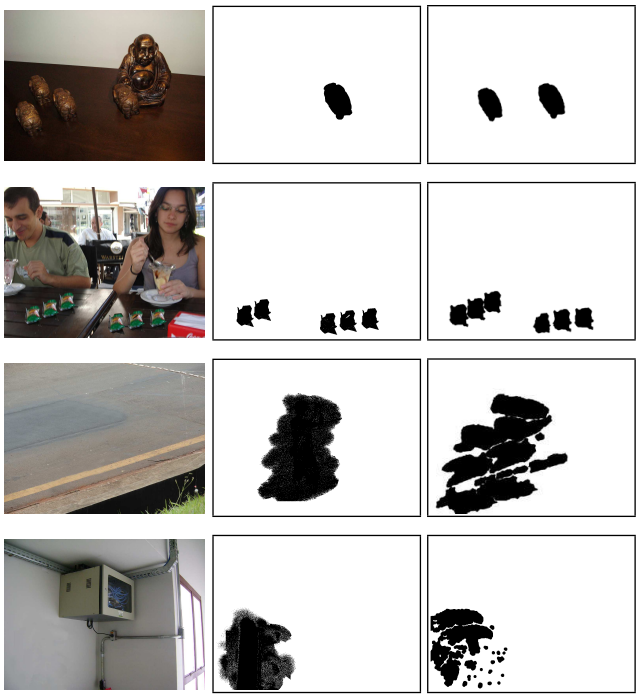

Fig. 3: Four training images with copy-move forgeries, their ground truth, and detection maps output by our method.

All matches obtained in perfectly flat areas, as in presence of saturation, are removed to reduce false alarms; likewise, very small regions are also deleted automatically through morphological filtering. Eventually, after elimination of unsuitable candidates, the image is classified as fake if at least one duplicated region is detected. To find also rotated copy-moves, we simply repeat the procedure for a number of rotations of the image, taking advantage of PatchMatch speed. Our experiments showed that a sampling step of 15 degrees guarantees accurate detection.

Fig.3 shows three images with copy-move forgeries, the corresponding ground truth, and the detection map output by our method. Note that the forgery is easily detected, and the map is quite accurate, even when original and copied regions are partially overlapping.

4. DECISION FUSION AND RESULTS

We implemented three forgery detection tools based on quite different approaches, local descriptors, block matching, and sensor noise. The latter tool, however, was discarded right away due to its poor detection performance. On the contrary, the local descriptor tool guarantees already an excellent performance on the training set, with a missing detection rate of 7.10%, and a false alarm rate of 2.29%.

It might seem difficult to improve upon such results through the fusion with a relatively weak copy-move detector. For sure, the latter cannot reduce the overall false-alarm rate, since its "pristine" decision means only that there is (probably) no copy-move forgery, but a splicing could still be present, so is basically useless. However, it can help reducing the missing-detection rate, by revealing all those copy-move forgeries that have escaped the previous detector, very likely because they are too small to impact on the descriptor. To this end, however, it is necessary that it be extremely specific, assuring that its "fake" decision is very reliable, and no new false alarm is introduced. Based on these considerations, we fine-tuned the copy-move detector by setting a very low threshold, thus detecting basically only rigid-translation forgeries, with little tolerance for other forms of processing. By so doing, we were able to detect the large majority of the copy-move forgeries in the training set with only 5 false alarms out of 1050 pristine images. Given these premises, the fusion rule consists in a simple OR of the decisions: an image is declared fake whenever any of the tools does so, and pristine only if both tools agree on that. With this rule, the score on the training set raises from 0.9530 to 0.9738.

Turning to the test set, comprising a total of 5713 images including an unknown number of fakes, groups participating in the Challenge had the opportunity to receive a limited feedback by submitting their classification once a day. Scores were then computed on a randomized subset of the test set to avoid disclosing valuable information through the system. The results on the test set were consistently worse, by about 2-3 percent point, than those obtained by cross validation on the training set. Despite the randomness of the feedback procedure, this fact indicates clearly some mismatch between training and test set and, therefore, a likely plateau for performance. Indeed, our final score, computed now on the whole test set, was 0.9421 as opposed to the 0.9738 on the training set. Note that the score obtained by running individually the two approaches is 0.8130 for the copy-move detection and 0.9150 for the method based on local descriptor. The described strategy allowed us to rank first in phase 1 of the Challenge. Interestingly, the scores of the first four groups, shown in Tab.III, were very close to one another suggesting that the plateau mentioned above has been probably reached.

#	Leader	Team	Score
1	Luisa Verdoliva	grip	0.9421
2	Guanshuo Xu	havefun	0.9373
3	Xinqi Lin	hyrup	0.9346
4	Licong Chen	Chen	0.9323
5	Khosro Bahrami	Fake Bluster	0.8574
6	Dev Sh	ITD	0.8240

Table 3: Final ranking (first six teams) for phase 1 of the Challenge.

5. CONCLUSIONS

We feel there are quite a few lessons to learn from this experience.

Under a strictly technical point of view, exploring locally the statistical features in the images is arguably the state-of-the-art approach in forgery detection. In particular, our implementation, with the selection and fusion of several different models based on their estimated AUC, provides a very competitive performance. Nonetheless, the fusion of more tools can further improve performance, especially when the additional techniques are very specific and can address reliably some niche problems, such as the detection of small rigid copy-moves. The copy-move detection technique provided exactly this result, also thanks to the fast PatchMatch search algorithm.

Under a wider point of view, we believe that this Challenge, with its large corpus of images and well-defined performance evaluation protocols, represents an important step for the growth of this field. It is worth remembering that the Challenge website [1] remains open for all interested research groups, who can download the images and constantly submit new results.

6. REFERENCES

[1] http://ifc.recod.ic.unicamp.br/fc.website/index.py?sec=0.

[2] D. Cozzolino, F. Gargiulo, C. Sansone, and L. Verdoliva, "Multiple classifier systems for image forgery detection," *International Conference on Image Analysis and Processing (ICIAP)*, vol. 8157, pp. 259-268, 2013.

[3] M. Fontani, T. Bianchi, A. De Rosa, A. Piva and M. Barni, "A framework for decision fusion in image forensics based on Dempster-Shafer theory of evidence," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 593,607, apr. 2013.

[4] G. Chierchia, S. Parrilli, G. Poggi, C. Sansone, and L. Verdoliva, "On the influence of denoising in PRNU based forgery detection," in *Proceedings of the 2nd ACM workshop on Multimedia in Forensics, Security and Intelligence* pp. 117–122, 2010.

[5] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, "A Bayesian-MRF approach for PRNU-based image forgery detection," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 554-567, apr. 2014.

[6] G. Chierchia, D. Cozzolino, G. Poggi, C. Sansone, and L. Verdoliva, "Guided filtering for PRNU-based localization of small-size image forgeries," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6273-6276, 2014.

[7] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "Fingerprint liveness detection based on Weber local image descriptor," *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, pp. 46-50, 2013.

[8] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "Local contrast phase descriptor for fingerprint liveness detection," *Pattern Recognition*, in press, 2014.

[9] Y.Q. Shi, C. Chen, and G. Xuan, "Steganalysis versus splicing detection," *International Workshop on Digital Watermarking*, vol. 5041, pp. 158-172, 2008.

[10] Z. He, W. Lu, W. Sun, and J. Huang, "Digital image splicing detection based on Markov features in DCT and DWT domain," *Pattern Recognition*, vol. 45, pp. 4292–4299, 2012.

[11] X. Zhao, S. Wang, S. Li, J. Li and Q. Yuan, "Image splicing detection based on noncausal Markov model," *IEEE International Conference on Image Processing*, pp. 4462-4466, sep. 2013.

[12] W. Wang, J. Dong, and T. Tan, "Effective image splicing detection based on image chroma," *IEEE International Conference on Image Processing*, pp. 1257–1260, 2009.

[13] D. Zou, Y.Q. Shi, W. Su, and G.R. Xuan, "Steganalysis based on markov model of thresholded prediction-error image," *International Conference on Multimedia and Expo*, pp. 1365–1368, 2006.

[14] T. Pevný, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 215-224, june 2010.

[15] J. Fridrich, and J. Kodovský, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868-882, june 2012.

[16] http://www.ws.binghamton.edu/.

[17] J. Fridrich, and J. Kodovský, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432-444, april 2012.

[18] V. Christlein, C. Riess, J. Jordan and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1841–1854, 2012.

[19] A. Langille, and M. Gong, "An efficient match-based duplication detection algorithm," *Canadian Conf. on Computer and Robot Vision*, pp. 1-8, 2006.

[20] S. Bayram, H.T. Sencar and N. Memon, "A survey of copy-move forgery detection techniques," *IEEE Western New York Image Processing Workshop*, 2010.

[21] 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, dec. 2010.

[22] R. Davarzani, K. Yaghmaie, S. Mozaffari, M. Tapak, "Copy-move forgery detection using multiresolution local binary patterns," *Forensic Science International* vol. 231, pp.6172, 2013.

[23] C. Barnes, E. Shechtman, A. Finkelstein, and D.B. Goldman, "PatchMatch: a randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 28, no. 3, aug. 2009.

[24] http://gfx.cs.princeton.edu/pubs/Barnes_2009_PAR/.