Splicebuster: a new blind image splicing detector

Davide Cozzolino, Giovanni Poggi, Luisa Verdoliva
DIETI, University Federico II of Naples, Italy

***Abstract*—We propose a new feature-based algorithm to detect image splicings without any prior information. Local features are computed from the co-occurrence of image residuals and used to extract synthetic feature parameters. Splicing and host images are assumed to be characterized by different parameters. These are learned by the image itself through the expectation-maximization algorithm together with the segmentation in genuine and spliced parts. A supervised version of the algorithm is also proposed. Preliminary results on a wide range of test images are very encouraging, showing that a limited-size, but meaningful, learning set may be sufficient for reliable splicing localization.**

***Index Terms*—Image forensics, forgery detection and localization, local image descriptors, blind algorithm.**

I. Introduction

Images and videos account already for the biggest share of traffic and storage space over the internet, and this trend is only going to increase in the near future. As manipulating multimedia content becomes ever more widespread and easy, the interest for digital image forensics is rapidly growing. Image forensic tools must address a wide variety of specific goals, form establishing the authenticity of an image, to discovering the presence of a manipulation, its type, its location, and so on. Indeed, many different forms of manipulation exist like copy-moving parts of an image, covering objects through inpainting, retouching details, or inserting material taken from a different source (splicing). Such diverse scenarios call for specific approaches and techniques. For example, to find copy-moves one looks for near-duplicates in the image, while to find a splicing one must discovery anomalies with respect to a typical behavior. These anomalies may be macroscopic, related to illumination or perspective inconsistencies, but skilled attackers avoid easily these errors. To detect accurate forgeries, statistical signal analysis tools are necessary.

In the last decade, many techniques have been proposed for splicing detection and localization, which can be classified based on the amount of prior information they rely upon. When either the host camera or an arbitrary number of images taken from it are available, one can estimate the so-called camera fingerprint, or photo-response non-uniformity noise (PRNU) pattern [1]. Being unique for any camera sensor, it allows to reliably identify the source camera, and also to detect and localize possible manipulations [1], [2] provided they are not too small.

A step below in this prior information scale, one can know or estimate the color filter array (CFA) and the interpolation filter characterizing the camera model. Given these pieces of information, one can detect transitions between original and spliced regions, as already suggested back in 2005 [3]. Several effective algorithms are based on this simple idea, like [4] and [5]. In alternative, detection and localization may rely on the different intensity and properties of the noise introduced in the image by different camera sensors [6], [7].

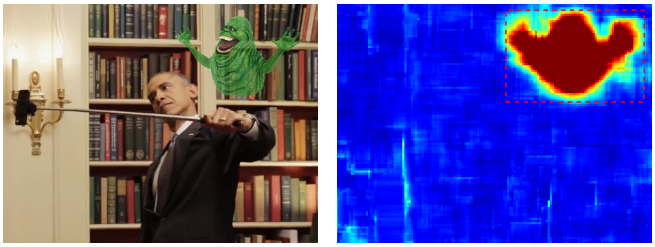

Fig. 1: Splicebuster working on a toy example. Local features extracted from the input image (left) are used to learn a model with two classes, associated with genuine and forged areas. The output heat map (right) indicates clearly a splicing in correspondence with the ghost.

A different form of prior information concerns the processing history of host image and splicing. In particular, assuming the images are always saved in compressed JPEG format, performing a splicing induces a double JPEG compression which leaves clear traces in the DCT coefficients of image blocks. Therefore, several methods have been proposed, like [8], [9] or [10], which exploit the statistical distribution of such coefficients.

All the above techniques rely on some strong and very specific hypothesis, which are not always met in practice. A more general approach consists in assuming that the different in-camera processing chain or out-camera processing history of host and splicing give rise to subtle differences in the high-pass content of the image. Whatever their origin, these patterns can be captured by some suitable features and classified by machine learning. In this context, the research focuses on the definition of the most expressive local features that account for such subtle differences. A first step in this direction dates back to 2004, with the model proposed in [11]. However, a major impulse comes only some years later with [12], where features based on both first-order and higher-order statistics of DCT coefficients are used, providing a performance gap with respect to the previous state of the art. In [13] the approach is extended to include also wavelet-based features, while [14] resorts to a noncausal Markov model. A local feature proposed originally for steganalysis [15], based on the co-occurrence of image residuals, is used in [16] for splicing detection with excellent results. In [17], the same features are used, but there is a switch from the machine learning paradigm to model based detection. Assuming that only genuine images are available,

978-1-4673-6802-5/15/$31.00 ©2015 IEEE

a model is learned for the host camera and used to detect data departing from the model. This latter work, therefore, borders the anomaly detection field, and also the camera model identification problem [18], [19].

Methods based on machine learning and feature modeling, though more general than the previous ones, have themselves a serious handicap, the need for a large training set. Sometimes, this set is simply not available. One may be given a single image and urged to decide whether it is pristine or forged, and which part of it has been manipulated. Barring fortunate cases, like copy-moves or double JPEG compression, this "blind" forgery detection problem may be very challenging.

In this paper we propose a new algorithm for the blind detection and localization of forgeries, nicknamed splicebuster. No prior knowledge is available on the host camera, on the splicing, or on their processing history. We use the co-occurrence based features proposed in [15] and, as in [17], follow an anomaly detection approach, learning a model for the features based on the very same image under analysis. In a first supervised scenario, the user is required to select a tentative training set to learn the model parameters, while in the unsupervised scenario, segmentation and model learning are pursued jointly by means of the expectation-maximization (EM) algorithm. Experimental results show that, despite the obvious loss of reliability due to the lack of an adequate training set, a very good performance can be obtained in most cases of interest.

II. Proposed method

To localize possible forgeries in the image we start from the approach proposed in [17], which is based on three major steps:

- defining an expressive feature that captures the traces left locally by in-camera processing;
- computing synthetic feature parameters (mean vector and covariance matrix) for the class of images under test, based on a suitable training set;
- using these statistics to discover where the features computed locally depart from the model, pointing to some possible image manipulation.

With respect to this paradigm, we have the major additional problem that no training set is available. A single image is given with no prior information. Still, we want to follow the same approach as before, computing model parameters and testing model fitting. This raises two distinct problems: *i)* even if an oracle told us which part of the image is pristine, the data available for training may be too scarce for reliable decision, and *ii)* we have no oracle, actually, so we must localize the forgery and estimate the parameters of interest at the same time. Indeed, if ideal single-image training does not provide reliable results, the whole approach is unsuitable for this task, no matter what we do. However, in Section 3, we will provide experimental evidence that single-image training is sufficient in most cases. Turning to the second issue, we will consider two scenarios, a supervised case, in which the user acts as an oracle, and an unsupervised case, where an EM-based procedure is used for simultaneous parameter estimation and image segmentation. These cases are explored in the following after describing the proposed feature.

A. Co-occurrence based local feature

Feature extraction is based on three main steps [15]

1) computation of residuals through high-pass filtering;
2) quantization of the residuals;
3) computation of a histogram of co-occurrences.

The final histogram is the feature vector associated with the whole image, which can be used for classification. To compute the residual image we use a linear high-pass filter of the third order, which assured us a good performance for both forgery detection [16], [17] and camera identification [19], defined as

$$r_{ij} = x_{i,j-1} - 3\,x_{i,j} + 3\,x_{i,j+1} - x_{i,j+2} \tag{1}$$

where x and r are origin and residual images, and i, j indicate spatial coordinates. The next step is to compute residual co-occurrences. To this end, residuals must be first quantized, using a very small number of bins to obtain a limited feature length. Therefore, we perform quantization and truncation as:

$$\widehat{r}_{ij} = \mathrm{trunc}_T(\mathrm{round}(r_{ij}/q)) \tag{2}$$

with q the quantization step and T the truncation value. We compute co-occurrence on four pixels in a row, that is

$$C(k_0,k_1,k_2,k_3) = \sum_{i,j} I(\widehat{r}_{i,j}=k_0, \widehat{r}_{i+1,j}=k_1, \widehat{r}_{i+2,j}=k_2, \widehat{r}_{i+3,j}=k_3)$$

where $I(A)$ is the indicator function of event A, equal to 1 if A holds and 0 otherwise. The homologous column-wise co-occurrences are pooled with the above based on symmetry considerations. Unlike in [15], we pass the normalized histograms through a square-root non-linearity, to obtain a final feature with unitary L2 norm. In fact, in various contexts, such as texture classification and image categorization, histogram comparison is performed by measures such as χ^2 or Hellinger that are found to work better than the Euclidean distance. After square rooting, the Euclidean distance between features is equivalent to do the Hellinger distance between the original histograms [20].

B. Supervised scenario

In this case, the user is assumed to take an active role in the process. She is required to select a bounding box, including the possible forgery, that will be subject to the analysis, while the rest of the image is used as training set (see Fig.1 for example). The analysis is carried out in sliding-window modality [17], using blocks of size $W \times W$, large enough to extract a meaningful feature, that is, the normalized histogram of co-occurrences, $\mathbf{h}$. The N blocks taken from the training area are used to estimate in advance mean and

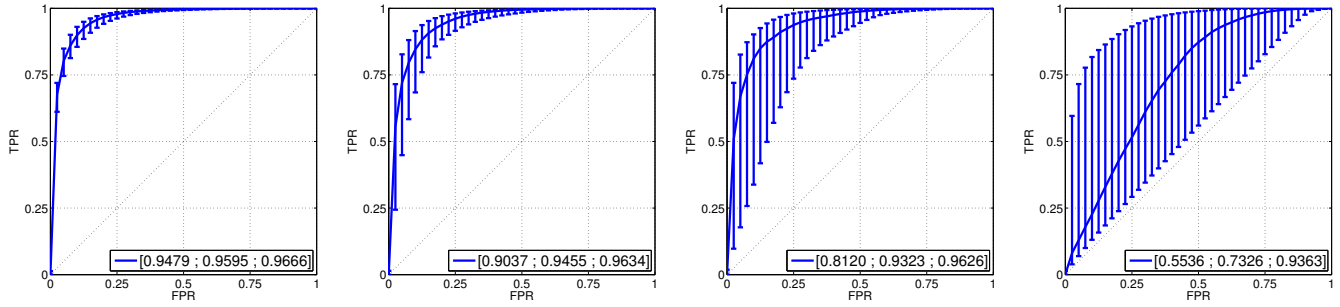

Fig. 2: Performance as a function of the training set size M: from left to right, M=50, M=10, M=5, M=1. For each FPR level, the bar ranges from the worst to the best TPR over the training sets. In parentheses, the worst, median and best AUC.

covariance of the feature vector

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{h}_n \tag{3}$$

$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{h}_n - \boldsymbol{\mu})(\mathbf{h}_n - \boldsymbol{\mu})^T \tag{4}$$

Then, for each block of the test area, the associated feature $\mathbf{h}'$ is extracted, and its Mahalanobis distance w.r.t. the reference feature $\boldsymbol{\mu}$ is computed

$$D(\mathbf{h}', \boldsymbol{\mu}; \boldsymbol{\Sigma}) = (\mathbf{h}' - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{h}' - \boldsymbol{\mu}) \tag{5}$$

Large distances indicate blocks that deviate significantly from the model. In the output map provided to the user, each block is given a color associated with the computed distance. Eventually, the user decides based on the visual inspection of the map (see again Fig.1).

Note that the user may repeat the process several times with different bounding boxes, implying that a meaningful analysis can be conducted even in the absence of any initial guess of the presence and location of a forgery.

C. Unsupervised scenario

In this case, after the feature extraction phase, carried out on the whole image with unit stride, we rely on an automatic algorithm to jointly compute the model parameters and the two-class image segmentation. Although there are many tools available for this task, for the time being, we resort to a simple expectation-maximization clustering.

As input, we need the mixture model of the data, namely, the number of classes, their probabilities, $\pi_0, \pi_1, \ldots$, and the probability model of each class. For us, the number of classes is always fixed to two, corresponding to the genuine area of the image (hypothesis $H0$) and the tampered area (hypothesis $H1$). We will consider two cases for the class models

1) both classes are modeled as multivariate Gaussian

$$p(\mathbf{h}) = \pi_0 \mathcal{N}(\mathbf{h}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \pi_1 \mathcal{N}(\mathbf{h}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$$

2) class H0 is modeled as Gaussian, while class H1 is modeled as Uniform over the feature domain Ω,

$$p(\mathbf{h}) = \pi_0 \mathcal{N}(\mathbf{h}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \pi_1 \alpha_1 I(\Omega)$$

We note explicitly that the Gaussian model is only a handy simplification, lacking more precise information on the feature distribution.

The first model is conceived for the case when the forged area is relatively large w.r.t. the whole image. Therefore, the two classes have the same dignity, and can be expected to emerge easily through the EM clustering. The block-wise decision statistic is the ratio between the two Mahalanobis distances.

When the forged region is very small, instead, the intra-class variability, mostly due to image content (*e.g.,* flat vs. textured areas) may become dominant w.r.t. inter-class differences, leading to wrong results. Therefore, we consider the Gaussian-Uniform model, which can be expected to deal better with these situations, and in fact has been often considered to account for the presence of outliers, *e.g.,* [21]. Note that, in this case, the decision test reduces to comparing the Mahalanobis distance from the Gaussian model with a threshold λ as already done in [17].

We do not choose between these two models, leaving the final say to the experimental analysis.

III. Experiments

We present now a number of experiments which provide insight into the potential of the blind techniques proposed here There is wide variety of manipulations of possible interest, and we have shown in [17] that the co-occurrence based feature allows one to detect and localize very well most of them. Here we focus only on splicing from other cameras and use 6 cameras of 6 different models and 4 manufacturers: Canon EOS 450D, Canon IXUS 95IS, Nikon D200, Nikon Coolpix S5100, Digimax 301, Sony DSC S780. For each camera we have a large number of images, which are cropped to size 768×1024 to speed-up processing.

Considering the limited training data available in this case, we must reduce as much as possible the feature length, so as to allow reliable estimates. Therefore, the truncation parameter is set to T=1, implying only three quantization levels for the residual, including 0. To balance losses, a relatively large quantization step, q=2 is used. Thanks to symmetries, the final feature has length 50, which is further reduced to 25 through PCA. The block size is 128×128, as a good compromise

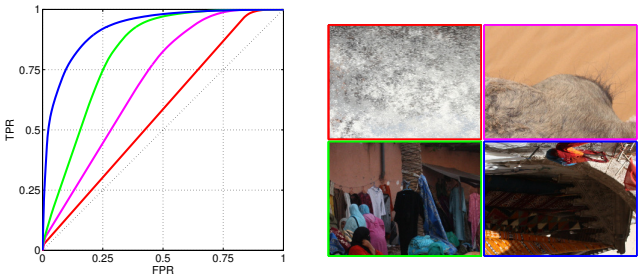

Fig. 3: Sample ROCs (left) obtained with single-image training and corresponding training images (right).

between accuracy and resolution. Since the results of the iterative EM algorithm depend strongly on the initialization we run it 30 times with different random initial parameters, selecting eventually the outcome for which the data exhibit the highest likelihood. Note that saturated and very dark areas tend to cause false alarms, and are hence excluded in this analysis.

A. *Dependence on training set size and quality*

Before showing results in the blind context, we carry out an experiment to study how results depend on the size and quality of the training set. We select a single camera as our host, and all the others as source of spliced material. The feature parameters for the host camera are estimated on a certain number M of training images. Then we test an equal number of genuine and fake blocks, deciding on their nature based on how their associated features fit with the camera model. Performance is measured in terms of true positive rate (TPR) vs. false positive rate (FPR). Notice that a very similar experiment was presented in [17], using always M=200. In Fig.2 we show the results obtained for M= 50, 10, 5 and 1, the latter amounting to single-image training. Since results may depend very much on the specific training images chosen, especially when just a few of them are used, the experiment is repeated several times with random instances of the training set, 200 times for the case M=1. In the figure, for each value of FPR, we show a bar going from the worst to the best TPR. The solid curve corresponds to median values.

It is clear that, with a large training set, say, 50 images, results are very good and depend very weakly on the specific set of images. With smaller sizes, 10 or 5, results are still generally good but present a larger variability. Going down to M=1, the dependence on the single training image becomes very strong. It is worth underlining, however, that for some instances of the single-image training the performance is quite good, not far from that of the 50-image case.

Fig.3 sheds some light on the nature of the good and bad training images. As could be expected, bad training images (red/magenta curves and boxes) are characterized by low contrast and limited variety of textures, sometimes highly unusual. On the contrary, good training images (green/blue curves and boxes) are quite varied, presenting bright and dark areas, with both textures and smooth contents. Considering that with such images performance is so good, one may argue

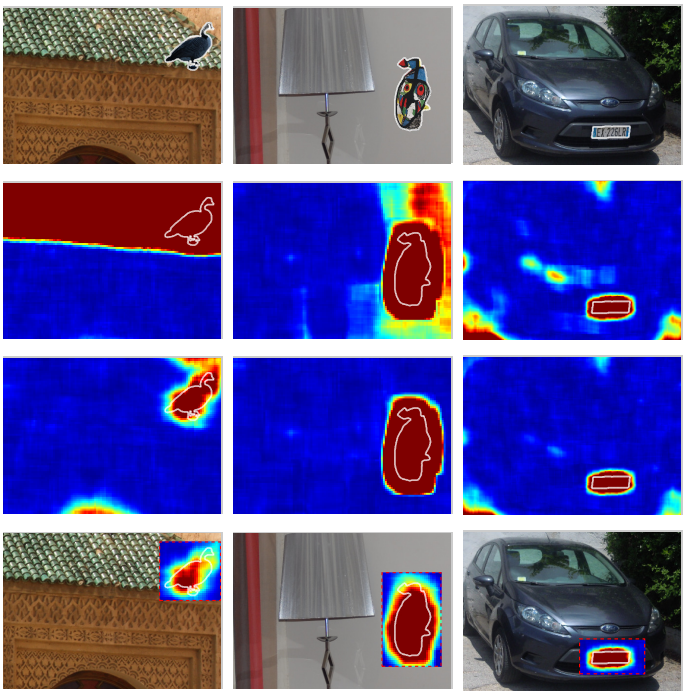

Fig. 4: Results for some selected examples. Top to bottom: forged images, maps obtained with the unsupervised method (GG and GU mixtures), and the supervised method.

that size is not really a limiting factor (at this level) provided sufficient variety is guaranteed. In addition, turning to our blind scenario, the training image is automatically well fit to the test, since most textures can be expected to be present in both sections.

B. *Analysis in controlled conditions*

To assess the performance of splicebuster we use visual inspection of results for some images with known splicing. In Fig.4 we show three selected examples, where the spliced area is highlighted, together with the maps provided by the variants of our method, that is, the unsupervised method with the two-Gaussian (GG) and Gaussian-Uniform (GU) mixture models (middle rows), and the supervised method (last row). The GU mixture provides always good results, while the GG mixture leads to some false alarms, a behavior observed also more in general. The supervised method is always very accurate. Note that the result of the unsupervised case can be used as a guide for the selection of the areas to investigate in more depth with the supervised approach.

C. *Comparison with the state of the art*

We now consider some comparisons with state of the art approaches. We used the 180 images coming from the Columbia Dataset[1]. Images are all in uncompressed formats with size from 757×568 to 1152×768. Spliced images were created using Adobe Photoshop with material coming from exactly two cameras, and no post processing was performed

[1] http://www.ee.columbia.edu/ln/dvmm/downloads/authsplcuncmp/.

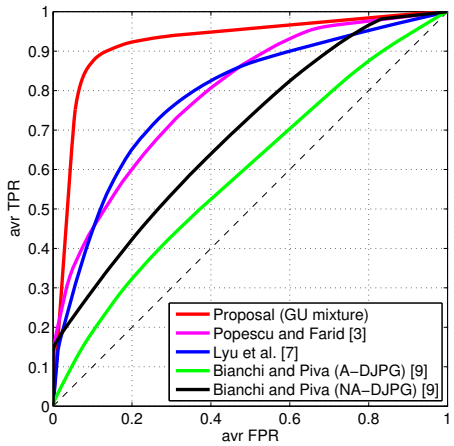

Fig. 5: Pixel-level ROCs on Columbia database.

[22]. We implemented the approaches of Popescu and Farid [3] based on CFA artifacts, and Lyu et al. [7] exploiting noise level inconsistencies. The code for the method of Bianchi and Piva [9] based on double JPEG compression was available online. Fig.5 shows ROCs obtained at pixel level and it can be seen that splicebuster performs much better than all references.

We also considered more realistic scenarios by using images publicly available on the net, where no information is provided about the nature of the splicings, hence it is possible that the images have undergone some post-processing operations. The first three are taken from the training set of the first IEEE Image Forensics Challenge[2], and come with a ground truth. The following four come from the test set of the same challenge, and the last two are drawn from the Worth1000 site[3]. In both cases no ground truth is available.

In Fig.6 next to each image, we show the heat maps obtained by the reference methods and the ones of the proposed approach in unsupervised (GU mixture) and supervised modality. In the latter case, we tested various bounding boxes. The visual inspection of the heat maps confirms the very good performance of splicebuster, except for some false alarms in the unsupervised case (dark blue areas correspond to saturated or very dark image regions and are not considered at all). Only in some cases, instead, the reference techniques provide sensible results, and the maps are typically less readable than those of the proposed method.

IV. Conclusion and Future Work

We proposed a new blind splicing detector. Results are definitely encouraging, especially if compared with reference methods. Still, there is much work ahead. Key parameters (like α in the GU mixture) are selected heuristically, for the time being. Likewise, the conversion from heat map to binary decision is still to perform. A major effort is then required to set up a sensible paradigm for objective performance assessment, and robustness to JPEG compression and other forms of post-processing should be explored.

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

[3]http://www.Worth1000.com

References

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

[2] 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, 2014.

[3] A.C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, 2005.

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

[5] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva, "Image Forgery Localization via Fine-Grained Analysis of CFA Artifacts," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1566–1577, october 2012.

[6] B. Mahdian and S. Saic, "Using noise inconsistencies for blind image forensics," *Image and Vision Computing*, vol. 27, no. 10, pp. 1497–1503, 2009.

[7] S. Lyu, X. Pan, and X. Zhang, "Exposing Region Splicing Forgeries with Blind Local Noise Estimation," *International Journal of Computer Vision*, vol. 110, no. 2, pp. 202–221, 2014.

[8] Y.-L. Chen and C.-T. Hsu, "Detecting Recompression of JPEG Images via Periodicity Analysis of Compression Artifacts for Tampering Detection," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 396–406, june 2011.

[9] T. Bianchi and A. Piva, "Image Forgery Localization via Block-Grained Analysis of JPEG Artifacts," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1003–1017, june 2012.

[10] I. Amerini, R. Becarelli, R. Caldelli, and A. Del Mastio, "Splicing forgeries localization through the use of first digit features," in *IEEE Workshop on Information Forensics and Security*, 2014, pp. 143–148.

[11] T.T. Ng and S.F. Chang, "A model for image splicing," in *IEEE International Conference on Image Processing*, 2004, pp. 1169–1172.

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

[13] 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.

[14] J. Zhao and W. Zha, "Passive forensics for region duplication image forgery based on harris feature points and local binary patterns," in *Mathematical Problems in Engineering*, 2013, pp. 1–12.

[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] D. Cozzolino, D. Gragnaniello, and L.Verdoliva, "Image forgery detection through residual-based local descriptors and block-matching," in *IEEE Conference on Image Processing (ICIP)*, October 2014, pp. 5297–5301.

[17] L. Verdoliva, D. Cozzolino, and G. Poggi, "A feature-based approach for image tampering detection and localization," in *IEEE Workshop on Information Forensics and Security*, 2014, pp. 149–154.

[18] M. Kirchner and T. Gloe, "Forensic camera model identification," in *Handbook of Digital Forensics of Multimedia Data and Devices*, T.S. Ho and S. Li, Eds. Wiley-IEEE Press, 2015.

[19] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "Evaluation of residual-based local features for camera model identification," in *International Workshop on Recent Advances in Digital Security: Biometrics and Forensics (BioFor)*, september 2015.

[20] R. Arandjelovic̀ and A. Zisserman, "Three things everyone should know to improve object retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2911–2918.

[21] A.C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 758–767, 2005.

[22] Y.F. Hsu and S.F. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency," in *IEEE International Conference on Multimedia and Expo*, 2006, pp. 549–552.

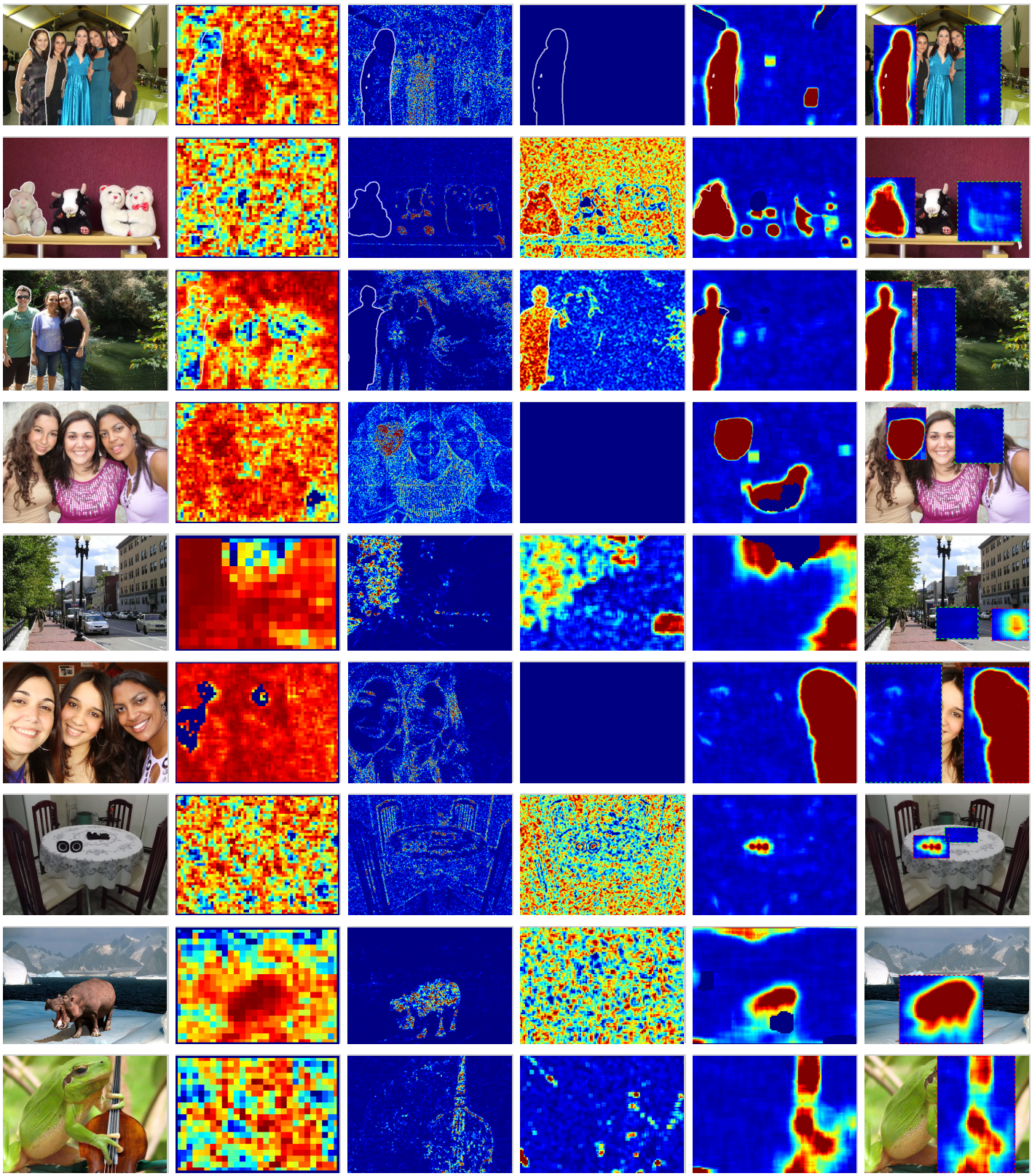

Fig. 6: Results of reference and proposed algorithms on some images available on the net. From left to right: forged image, heat maps obtained with the method of Popescu and Farid [3], Lyu et al. [7], Bianchi and Piva [9], splicebuster in unsupervised (GU mixture) and supervised modality.