Evaluation of Random Field Models in Multi-modal Unsupervised Tampering Localization

Paweł Korus*†, and Jiwu Huang*
*College of Information Engineering, Shenzhen University, Shenzhen, China
†Department of Telecommunications, AGH University of Science and Technology, Kraków, Poland
e-mail: pkorus@agh.edu.pl, jwhuang@szu.edu.cn

***Abstract*—While it is commonly known that successful forensic detectors should combine clues from various forensic features, unsupervised multi-modal tampering localization is still an open problem. State-of-the-art fusion methods perform simple pixel-wise combination of the input tampering maps. In this study, we show that pixel-wise combination is sub-optimal and successful fusion needs to model dependencies between neighboring pixels and exploit the content of the tampered image. We evaluate two methods based on conditional random fields and demonstrate that they can exploit image content and precisely delineate the shape of the forgery. In contrast to existing methods based on explicit image segmentation, such an approach does not suffer from subtle object removal forgeries where meaningful segments do not exist. We also demonstrate that existing performance measures are insufficient to accurately assess tampering localization performance. Further work in this direction is needed.**

***Index Terms*—digital image forensics; content forgery; tampering localization; decision fusion; random field models; photo-response non-uniformity; color-filter array**

I. Introduction

Reliable forensic analysis should incorporate multiple detectors sensitive to diverse traces of prospective forgeries. Individual detectors' responses can be combined into a single, more reliable decision. In addition to classical methods, like majority voting, the problem can also be addressed with sophisticated frameworks based on the Dempster-Shafer theory of evidence (DSTE) [1] or fuzzy logic [2]. Such frameworks deal with uncertainty and compatibility of candidate decisions in a systematic way by defining tables of the expected (and unexpected) combinations of the traces. Use of supervised learning techniques is discouraged due to poor scalability.

Forensic decision fusion has been studied in depth only for tampering detection, and its extension to unsupervised tampering localization is still an open problem. State-of-the-art tampering map fusion methods are based on pixel-wise application of simple combination rules [3, 4]. A recent evaluation found that naive map summation and product fusion deliver the best performance [5]. Early results based on the DSTE framework are available in [6], but still rely on pixel-wise application of this combination rule.

The diversity of forensic traces / detectors clearly suggests that a naive pixel-wise approach is suboptimal. The problem can be illustrated (Fig. 1) by comparing the output of two popular detectors based on extremely different analysis windows: a CFA detector [7] (with 8×8 px window), and a PRNU detector [8] (128×128 px). Large-window detectors yield coarse shapes and no pixel-wise combination rule can fuse them optimally with a small-window detector. For best performance, the detected regions should be cross-referenced with objects from the image - an inevitable next step of a human forensic analyst.

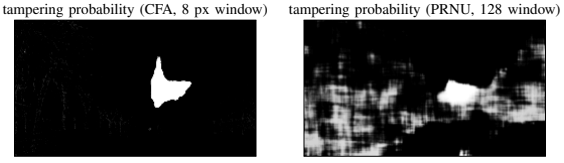


Fig. 1. Detected object shape mismatch in tampering probability maps from forensic detectors operating on different scales of analysis.

Exploitation of image content to guide tampering localization is a still an open problem. The first attempts extracted forensic features from image segments instead of typically used square windows [9]. Assuming reliable manual segmentation is available, it becomes possible to detect even small forgeries with detectors that typically require large analysis windows, e.g., the PRNU detector [10]. However, such an approach is difficult to use in practice - in particular in an unsupervised setting. The main challenges involve not only proposition of semantically meaningful objects in the image but also handling of subtle object removal forgeries with no clear object boundaries. A recently proposed solution involves the use of guided image filtering [11] which weights samples in the correlation statistics by perceptual similarity.

In this study, we evaluate another approach based on conditional random fields (CRF) with content-dependent interaction potentials. Analogous adaptive terms are used in state-of-the-art image segmentation schemes [12]. We will demonstrate that such an approach can accurately delineate even complex shapes of inserted objects, and does not suffer from subtle forgeries involving object removal. Hence, it solves one of the main challenges in the exploitation of image content in unsupervised tampering localization. We will also demonstrate that due to the limitations of existing performance assessment protocols, the benefits of such methods are difficult to measure reliably, indicating the need for future research.

Our evaluation addresses the problem of multi-modal decision fusion in an unsupervised localization setting. We assess the localization performance on high-quality realistic forgeries,

978-1-5090-1138-4/16/$31.00

and include two popular detectors operating on different scales. We show that adoption of CRF-based models warrants a significant performance improvement over existing pixel-wise approaches with additional benefits of straightforward image content exploitation.

This paper is organized as follows. In Section II we briefly introduce the adopted forensic detectors. In Section III we describe the considered decision fusion methods and the structure of the random fields. Our experimental evaluation is presented in detail in Section IV. We conclude in Section V. Supplementary materials and our dataset are available online: http://kt.agh.edu.pl/~korus/publications/2016-wifs.

II. Tampering Localization Methods

In this section, we briefly introduce the considered CFA [7] and PRNU detectors [8]. We configured both methods to yield real-valued tampering probability maps of the same size (resolution of 8×8 px blocks).

A. Detection of Color Filter Array Artifacts

Virtually all consumer-level digital cameras use a monochrome imaging sensor. Capturing color images requires a color filter array in front of the sensor and a *demosaicing* step to produce full-frame RGB images. As a result, an authentic digital photograph is expected to exhibit consistent periodic interpolation artifacts. The considered detector [7] compares prediction error within interpolated pixels against acquired pixels. It operates on small, non-overlapping image blocks (8×8 px for best performance). In order to improve elimination of false positives, we reset the scores of saturated blocks.

B. Detection of Photo-Response Non-Uniformity Artifacts

Imperfections of the imaging sensor's manufacturing process introduce minor variations in the sensitivity of individual pixels. Inconsistencies of this noise in different image regions indicate potential tampering. The detector computes correlation of a previously obtained noise signature with its estimate from the investigated image [8]. In tampered areas, the expected correlation is zero; in authentic ones, the expected correlation changes depending on the image content. A correlation predictor uses local image features to estimate the expected correlation, and then the detector can compute the probability of tampering based on Bayesian analysis (see [13] for a detailed description and source code).

For reliable statistics, validation of the PRNU signature requires large analysis windows (e.g., 128×128 px [8]). Hence, the windows are overlapping, and the obtained tampering probability is attributed to the central pixel only. In our work, we can use a different detector to eliminate prospective false positives, and therefore use analysis windows of size 64×64 px to minimize mismatch between the detectors.

III. Decision Fusion Methods

The goal of a fusion procedure is to generate a single binary tampering map based on multiple candidate tampering probability maps (possibly by exploiting the image content $\mathbf{I}$):

$$\left(\{\mathbf{c}^{(d)} \in [0,1]^N : d \in \mathcal{D}\}, \mathbf{I}\right) \rightarrow \mathbf{t} \in \{0,1\}^N \tag{1}$$

where $\mathcal{D}$ is a set of available detectors; in our study $\mathcal{D} = \{\text{cfa}, \text{prnu}\}$. We assume that the input maps are of the same size $N = w \times h$. For notation simplicity, we address individual pixels with a single index (i.e., t_i denotes the i-th pixel of $\mathbf{t}$).

A. Pixel-wise Combination Rules

The most straightforward approach fuses the input maps by pixel-wise application of a combination rule. A recent study concluded that naive map summation and product fusion yield the best results [5]. Therefore, we consider the following combination rules as a baseline for performance comparison: *sum*, *product*, and *disjunction*. Let τ denote the decision threshold and the tilde operator denote $\tilde{x} = 1 - x$. The *sum* fusion resolves to:

$$t_i = \left(c_i^{(\text{cfa})} + c_i^{(\text{prnu})}\right)/2 > \tau \ . \tag{2}$$

The *product* fusion is defined by:

$$t_i = \left(c_i^{(\text{cfa})} c_i^{(\text{prnu})}\right) \left(c_i^{(\text{cfa})} c_i^{(\text{prnu})} + \tilde{c}_i^{(\text{cfa})} \tilde{c}_i^{(\text{prnu})}\right)^{-1} > \tau \ . \tag{3}$$

The *disjunction* fusion is defined by:

$$t_i = \left(c_i^{(\text{cfa})} > \tau\right) \vee \left(c_i^{(\text{prnu})} > \tau\right) \ . \tag{4}$$

Additionally, we consider an *empirical* fusion method where the combination rule is learned from data. We quantize candidate tampering probabilities into 8 bins, and count occurrences of positive & negative ground truth labels within each bin. Hence, the rule is defined by a 8×8 array with empirical probabilities; the array is averaged with its transposed version for symmetry, and interpolated by local regression smoothing for generalization to arbitrary real-valued inputs.

For all pixel-wise combination rules, we post-process the resulting binary maps with heuristic cleaning. We used morphological opening with a disk-shaped structural element (SE) of size 15×15 px. For the *disjunction* rule we consider an additional variant of the rule which exploits the possibility of customizing heuristic cleaning for different detectors. The variants are referred to as *common cleaning* (CC) and *independent cleaning* (IC), respectively. In the latter, the PRNU map is cleaned with a larger SE (31×31 px - quarter of the window size) and the eroded shape of the detected region is recovered by morphological dilation (19×19 px SE).

B. Modeling Neighborhood Interactions

Conditional random fields constitute a popular approach to modeling neighborhood interactions in computer vision and image processing [14]. In such a model, the decisions for individual pixels are no longer independent, but take into account the decisions for their neighbors. Recently, a Markov random field has been used to improve the localization performance of the PRNU detector [15]. Here, we consider extended models where the potentials are constructed in a content-dependent way [12, 13].

The localization problems resolves to finding a labeling $\mathbf{t}$ of image blocks that minimizes the following energy function:

$$E(\mathbf{t}) = \frac{1}{|D|} \sum_{d \in \mathcal{D}} \sum_{i=1}^{N} \psi_\tau(c_i^{(d)}|t_i) + \sum_{i=1}^{N} \sum_{j \in \Xi_i} \phi_p(t_i, t_j) \quad (5)$$

where ψ_τ, and ϕ_p denote the unary and pairwise potentials, respectively, and Ξ_i denotes the neighborhood of pixel i. We use the approach from [16] to construct the unary potentials:

$$\psi_\tau(c|t) = -\log \max\left(\Psi_{\min}, \Psi_\tau(c|t)\right), \quad (6)$$

with $\Psi_{\min} \in [0, 1]$ and:

$$\Psi_\tau(c|t) = \begin{cases} 1 - \frac{c}{2\tau} & \text{for } t = 0, \\ 1 + \frac{c}{2(1-\tau)} - \frac{1}{2(1-\tau)} & \text{for } t = 1, \end{cases} \quad (7)$$

where $\tau \in (0, 1)$ is a quasi-threshold that equalizes potentials for both decisions, i.e., $\psi_\tau(\tau, 0) = \psi_\tau(\tau, 1)$. Setting a minimal value $\Psi_{\min}$ (0.001 in our experiments) prevents the nodes from becoming fixed to certain decisions (due to infinite energy).

The pairwise potentials penalize differences in decisions among neighboring pixels. We consider two versions of neighborhood interactions. In a *grid* CRF we use 8-nearest neighbors as the pixels' neighborhood Ξ_i and pairwise potentials of the form $\phi_p(t_i, t_j) = \beta_{ij}|t_i - t_j|$ where:

$$\beta_{ij} = \beta_0 + \beta_1 \text{exp}\left(-\frac{|I_i - I_j|^2}{2\theta_2^2}\right). \quad (8)$$

The parameters β_0 and β_1 denote the default and the content-dependent interaction strengths. The latter fades off exponentially with L2 distance between pixels' RGB vectors I_i and I_j. The desired pixel similarity is controlled with parameter θ_2. We also consider a *dense* CRF spanned over a fully connected grid, i.e., with $\Xi_i = \{1, \ldots, N\} \setminus i$, which uses Gaussian potentials of the form $\phi_p(t_i, t_j) = \beta_{ij}|t_i - t_j|$ where:

$$\beta_{ij} = \beta_0 \text{exp}\left(-\frac{|p_i - p_j|^2}{2\theta_0^2}\right) + \quad (9a)$$

$$+ \beta_1 \text{exp}\left(-\frac{|p_i - p_j|^2}{2\theta_1^2} - \frac{|I_i - I_j|^2}{2\theta_2^2}\right) \quad (9b)$$

where p_i denotes a vector of (x, y) coordinates of the i-th pixel on the image grid. In both models, the neighborhood interaction terms encourage nearby similar pixel to assume the same label. Hence, the CRF-based fusion methods do not require heuristic map post-processing.

IV. Experimental Evaluation

Our evaluation involves tampering localization for high-quality realistic forgeries. Using modern photo editing software, we manually created a dataset of 120 diverse forgeries involving not only object insertion and removal, but also more subtle changes to the image content, not necessarily detectable by the considered detectors (e.g., shadows or reflections of inserted objects). The images originate from 3 cameras: Sony α57 (personal photo collection), and Nikon D90, and D7000 (RAISE dataset [17]). We used TIFF images (directly acquired from [17] or converted with *dcraw*) of size 1920×1080 px (cropped from the middle). The PRNU was obtained with the standard MLE estimator [18] from 200 favorable natural images (Nikon) or from 90 dedicated out-of-focus flat images (Sony). The predictor was trained on 25,000 patches from 50 diverse images.

TABLE I
Empirically chosen parameters for the CRFs.

	β_0	β_1	θ_0	θ_1	θ_2
grid	0.25	2.00	-	-	25
dense	3.75	3.00	3.5	10	20

Finally, we analyzed the tampered images with the CFA and PRNU detectors (Section II). Having only 2 decision labels, and guaranteed sub-modular potentials ($\beta_0, \beta_1 \geq 0$), we used a graph cuts-based solver [19] from the UGM toolbox [20] to quickly find the optimal tampering map for the *grid* CRF. For the *dense* CRF we used a recently proposed efficient solver based on iterative mean-field approximations [21].

A. Parameter Selection & Evaluation Protocol

The *grid* CRF is controlled by 3 parameters $(\beta_0, \beta_1, \theta_2)$ and the *dense* CRF by 5 parameters $(\beta_0, \beta_1, \theta_0, \theta_1, \theta_2)$. The parameters have a sound interpretation and reasonable values may be chosen empirically by trial and error on a few representative images (Table I). However, in order to assess the sensitivity of the localization performance, we also consider an automated procedure that maximizes the average F_1 score:

$$F_1 = \frac{2 \cdot tp}{2 \cdot tp + fn + fp}, \quad (10)$$

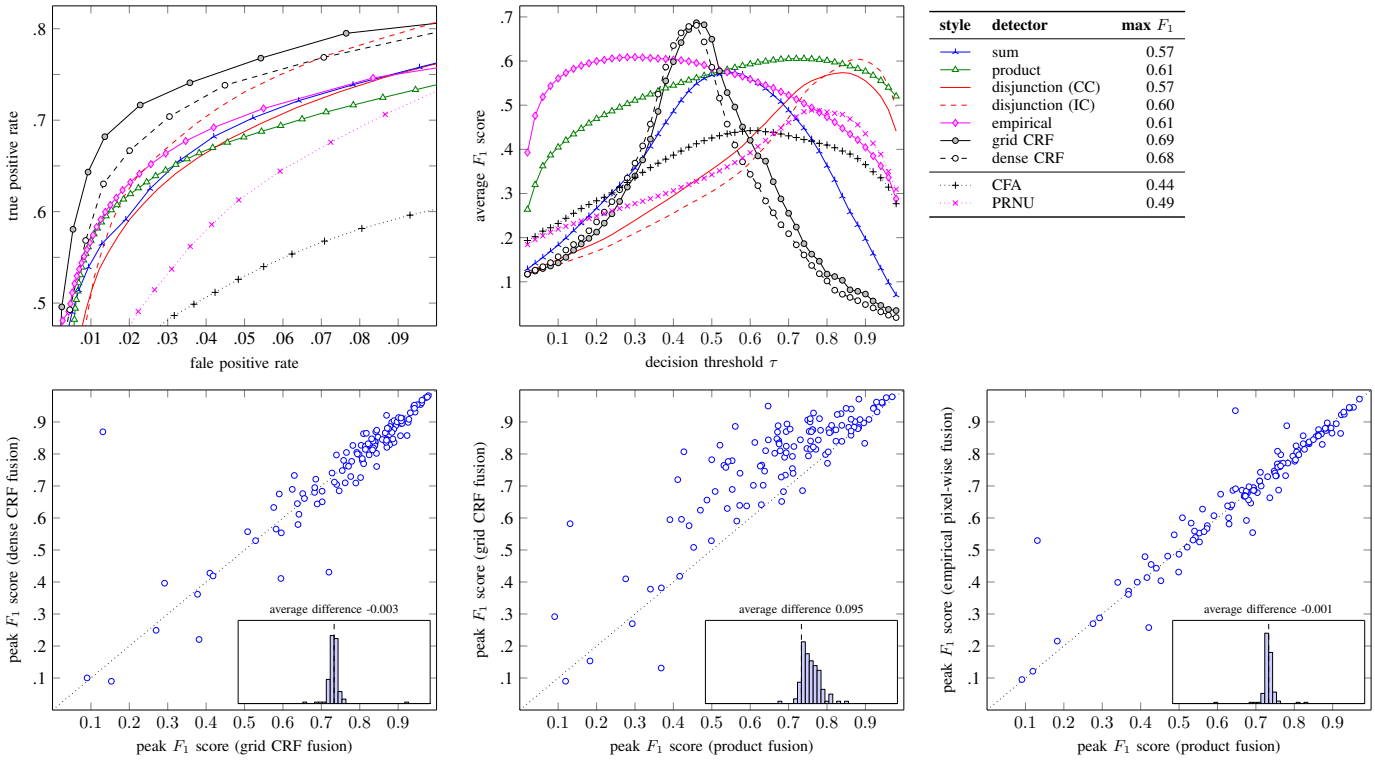
where tp, fn, fp denote the observed true positives, false negatives, and false positives. Due to limited number of test images, we resort to cross-validation with random-sampling:

1) Split the test set randomly into 25% : 75% for training and testing, respectively.
2) Use random parameter sweep (for the dense CRF) or grid search (grid CRF) to find the parameters maximizing the average F_1 for the training set.
3) Test the best parameters on the remaining images.
4) Repeat steps 1-3 for M random splits of the dataset.

In random parameter sampling we used 1,000 attempts with $\beta \leftarrow \mathcal{U}(0, 10)$ and $\theta \leftarrow \mathcal{U}(0.1, 50)$. In grid search, we uniformly sampled 256 points from $[0, 3] \times [0, 3]$ for β and kept the empirically chosen θ for simplicity. We used $M = 10$ for the dense CRF and $M = 30$ for the grid CRF. When testing, we sweep the threshold over 49 values uniformly chosen from $(0, 1)$. To speed up training, we used 19 thresholds and approximated the curve by cubic interpolation.

B. Empirical Evaluation Results

The obtained tampering localization results are collected in Fig. 2. Both the ROC curves (top left) and the average F_1 scores (top right) clearly show significant improvement of the decision fusion methods over individual detectors, and of the CRF-based fusion over standard pixel-wise methods.



style	detector	max F_1
	sum	0.57
	product	0.61
	disjunction (CC)	0.57
	disjunction (IC)	0.60
	empirical	0.61
	grid CRF	0.69
	dense CRF	0.68
	CFA	0.44
	PRNU	0.49

Fig. 2. Tampering localization performance for the considered decision fusion methods with empirically chosen parameters of the CRFs: the receiver operation characteristics (top left); the average F_1 score vs. decision threshold τ (top right); and comparison of peak localization performance for individual images: grid CRF vs. dense CRF (bottom left); grid CRF vs. pixel-wise product fusion (bottom middle); empirical vs. product pixel-wise fusion (bottom right).

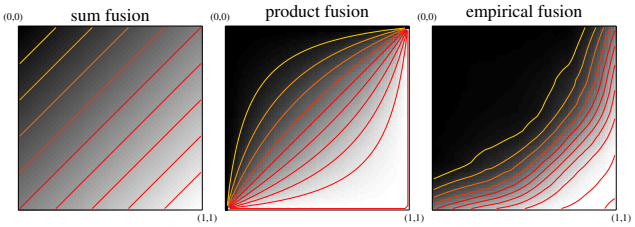


Fig. 3. Visualization of combination rules; the xy axes correspond to the input candidate scores; brightness corresponds to the output score.

Apart from somewhat inferior results of the *sum* and *disjunction (CC)* fusion, we observed similar performance for most pixel-wise methods. The *disjunction* rule could further benefit from individualized post-processing - especially, when larger false alarm rates are acceptable. Overall, the best pixel-wise performance was obtained by the *empirical* and *product* fusion, with slight advantage of the former. Although the measurable improvement (both F_1 and accuracy-wise) was negligible, we observed important qualitative differences. By examining contours of the combination rule (Fig. 3) we can clearly see that the product fusion is in fact not an accurate model. Despite certain shape similarities to the empirical fusion near (1,1), the rules diverge as the candidate scores decrease. When any of the candidate scores is zeroed, the product fusion always deems the pixel authentic which is clearly a bad assumption. Nevertheless, the performance penalty for

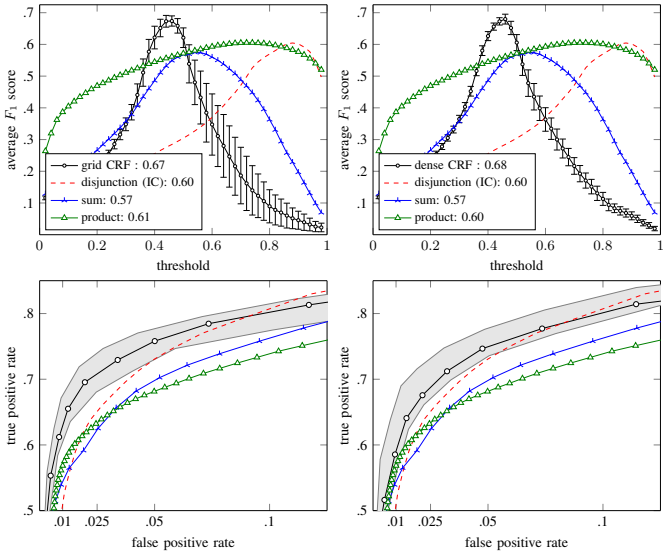


Fig. 4. Localization performance stability for CRF-based fusion: error bars correspond to standard deviation of the F_1 scores; shaded areas corresponds to the worst and the best observed ROC curves.

this violation seems to be negligible in practice since zeroed values are rare. Finally, we observed that the *empirical* fusion provides a certain visual advantage - uncertain regions are better attenuated, which is also reflected in the maximum of $F_1(\tau)$ shifted towards lower τ.

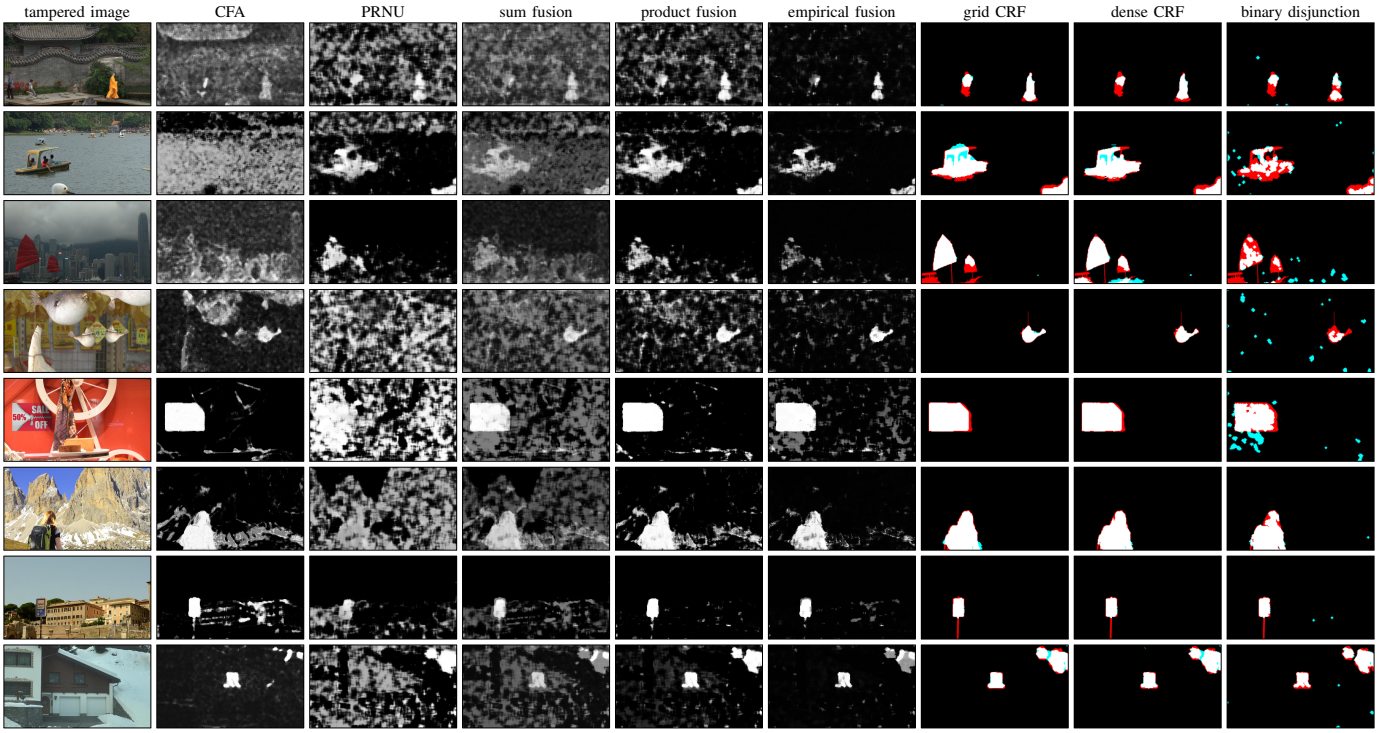


Fig. 5. Example tampering localization results; binary decision maps correspond to the best map (F_1 score-wise) over all possible choices of the decision threshold and are color coded: true positive (white), false negative (red), false positive (cyan), true negative (black).

While incorporation of neighborhood interactions yields significant performance improvement over conventional pixel-wise methods, we did not observe major differences between the CRF models. While we obtained slightly better ROC curves for the grid CRF, the F_1 evaluation shows no difference between the models. More detailed analysis of individual cases also revealed no improvement. The scatter-plot of individual per-image peak F_1 scores[1] (bottom left plot in Fig. 2) confirms that there are no clear advantages of either of the methods. This can be contrasted with clear benefits of the grid CRF over *product* fusion (bottom middle plot in Fig. 2).

Performance stability analysis also reveals no significant differences between the CRF models (Fig. 4). We followed the cross-validation procedure described in Section IV-A. The obtained results are nearly identical as for the hand-picked parameters, and show stable improvement over pixel-wise methods, and minor ROC advantage of the simpler *grid* CRF.

Example tampering localization results are shown in Fig. 5. For the *product*, *sum*, and *empirical* fusion we show real-valued localization maps. For the remaining fusion methods, we show color-coded binary decision maps, corresponding to the best observed localization (F_1 score-wise) over possible choices of the decision threshold τ. It can be observed that CRF-based methods yield the best results with superior shape representation, even for small details of the inserted objects.

[1]We define *peak* F_1 *scores* as the maximal F_1 score for each test image (over all possible decision thresholds τ).

C. Limitations of Localization Performance Measures

Visual inspection of individual forgeries clearly demonstrates the efficiency of the adopted content-dependent CRFs. Boundaries of inserted objects are closely followed, and even small details of their shapes are accurately detected. At the same time, the method does not suffer in the presence of more subtle object removal forgeries. Fig. 6 illustrates this for two example forgeries and compares the results of localization guided by the inserted object (1st column) and by the background only (2nd column). The latter map is similar to a simpler grid CRF without content-dependent interactions[2] ($\beta_0 = 0, \beta_1 = 1.8$, 3rd column). For reference, the results of pixel-wise product fusion are shown in 4th column.

Despite significant differences in shape representation, we observed that currently used measures for evaluation of tampering localization schemes deliver inadequate results. Adoption of content dependencies often introduces either minimal improvement (bottom row) or even significant deterioration (top row) of the measurable performance (both in terms of F_1 scores and accuracy A) despite clearly more precise detection. This is particularly well visible on small details of the tampered shapes, e.g., the roof or the back of the boat, which are easily removed by conventional content-independent neighborhood interactions. In part, the problem can be attributed to *collateral damage*, i.e., pixels with different values

[2] We repeated our evaluation with a simplified grid CRF without content dependencies. Quantitatively, we obtained nearly the same results as with adaptive interactions - both in terms of ROC and F_1 performance.

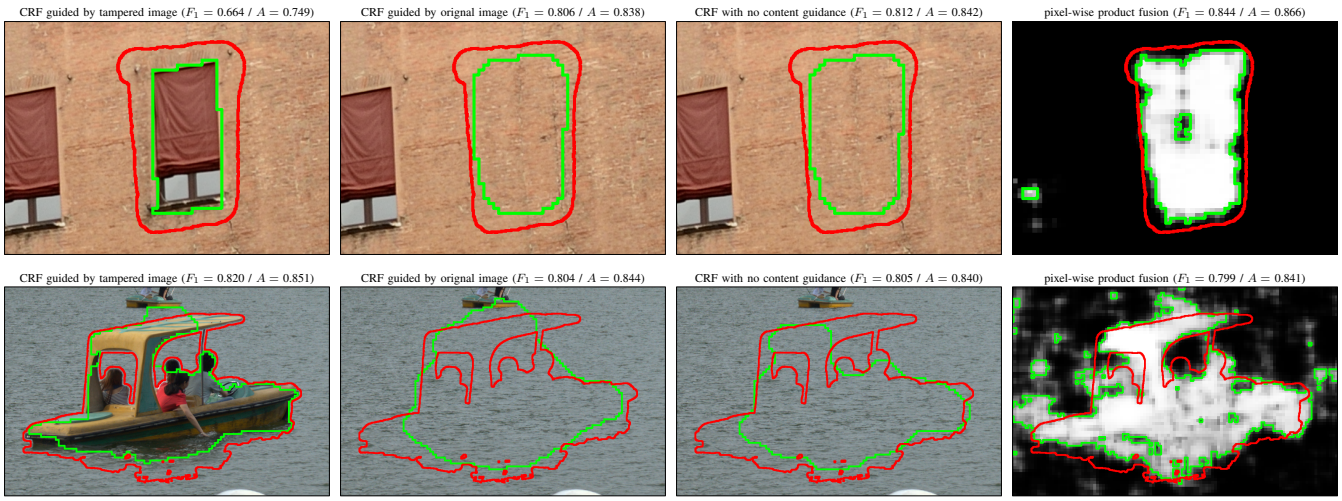


Fig. 6. Impact of content-dependency for two example forgeries (grid CRF): note significant deterioration of measurable localization quality (F_1 score / accuracy A) in the first example, and infinitesimal improvement in the second when the inserted object is used for guidance; note also similar results when no content guidance is used (3rd column) and when background content is used for guidance (2nd column); Legend: detected region (green); ground truth (red).

but the same semantic content (e.g., additional border around the inserted objects caused by image blending). This clearly shows the need for more accurate evaluation metrics.

V. Conclusions

In this study, we focused on fusion of diverse tampering localization maps, obtained from forensic detectors operating on different scales of analysis. We used conditional random fields to model dependencies between neighboring image blocks in a content-adaptive way. Such an approach significantly improves the detection of complex shapes, and - in contrast to existing methods based on image segmentation - does not suffer from subtle object removal forgeries.

We have also shown that existing metrics for evaluation of tampering localization performance are inadequate. They ignore spatial relationships of detected objects, and are not accurate enough to reflect differences between rough and precise object representation. This clearly shows the need for further research towards better evaluation protocols.

While our framework supports an arbitrary number of detectors, it expects unified tampering probability maps as an input. Further work is required to include arbitrary detectors, possibly operating on different domains (e.g., segments or super-pixels) or incapable of probabilistic output.

Acknowledgment

The research leading to these results was partly supported by NSFC (61572329, 61402295), Shenzhen R&D Program (GJHZ20140418191518323, JCYJ20160328144421330), and Guangdong NSF (2014A030313557).

References

[1] M. Fontani et al., "A framework for decision fusion in image forensics based on Dempster-Shafer theory of evidence," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 4, 2013.

[2] M. Barni and A. Costanzo, "Dealing with uncertainty in image forensics: a fuzzy approach," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2012.

[3] L. Gaborini et al., "Multi-clue image tampering localization," in *Proc. of IEEE Int. Workshop on Inf. Forensics and Security*, 2014.

[4] D. Cozzolino et al., "Image forgery localization through the fusion of camera-based, feature-based and pixel-based techniques," in *Proc. of IEEE Int. Conf. on Image Processing*, 2014.

[5] D. Cozzolino et al., "Multiple classifier systems for image forgery detection," in *Image Analysis and Processing*, vol. 8157 of *LNCS*. 2013.

[6] P. Ferrara et al., "Unsupervised fusion for forgery localization exploiting background information," in *Proc. of IEEE Int. Conf. on Multimedia & Expo Workshops*, 2015.

[7] P. Ferrara et al., "Image forgery localization via fine-grained analysis of cfa artifacts.," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, 2012.

[8] M. Chen et al., "Determining image origin and integrity using sensor noise," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, 2008.

[9] M. Barni et al., "Identification of cut & paste tampering by means of double-jpeg detection and image segmentation," in *Proc. of IEEE Int. Symposium on Circuits and Systems*, 2010.

[10] G. Chierchia et al., "PRNU-based detection of small-size image forgeries," in *Proc. of Int. Conf. on Digital Signal Processing*, 2011.

[11] G. Chierchia et al., "Guided filtering for PRNU-based localization of small-size image forgeries," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2014.

[12] M.-M. Cheng et al., "DenseCut: Densely connected CRFs for realtime GrabCut," *Computer Graphics Forum*, vol. 34, no. 7, 2015.

[13] P. Korus and J. Huang, "Multi-scale analysis strategies in PRNU-based tampering localization," *IEEE Trans. Inf. Forensics Security*, To appear.

[14] S. Z. Li, *Markov Random Field Modeling in Image Anlaysis*, Springer-Verlang, 2001.

[15] G. Chierchia et al., "A Bayesian-MRF approach for PRNU-based image forgery detection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 4, 2014.

[16] P. Korus and J. Huang, "Multi-scale fusion for improved localization of malicious tampering in digital images," *IEEE Trans. Image Process.*, vol. 25, no. 3, 2016.

[17] D. T. Dang-Nguyen et al., "RAISE - a raw images dataset for digital image forensics," in *Proc. of ACM Multimedia Systems*, 2015.

[18] "Dde laboratory," http://dde.binghamton.edu/, visited in Sept. 2015.

[19] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, 2001.

[20] M. Schmidt, "UGM: A matlab toolbox for probabilistic undirected graphical models," http://www.cs.ubc.ca/~schmidtm/Software/UGM.html, 2011.

[21] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. of NIPS*, 2011.