Improved Tampering Localization in Digital Image Forensics based on Maximal Entropy Random Walk

Paweł Korus, *Member, IEEE,* Jiwu Huang, *Fellow, IEEE*

***Abstract*—In this paper we propose to use maximal entropy random walk on a graph for tampering localization in digital image forensics. Our approach serves as an additional post-processing step after conventional sliding-window analysis with a forensic detector. Strong localization property of this random walk will highlight important regions and attenuate the background - even for noisy response maps. Our evaluation shows that the proposed method can significantly outperform both the commonly used threshold-based decision, and the recently proposed optimization-based approach with a Markovian prior.**

***Index Terms*—digital image forensics; tampering localization; visual saliency; maximal entropy random walk; first-digit features; JPEG splicing**

I. Introduction

Precise localization of tampered image regions is one of the most challenging problems in digital image forensics. While many forensic features are known to differ between pristine and tampered images, their application for blind localization poses additional challenges. Typically, the localization capability is obtained by analyzing the image in a sliding-window manner. Existing schemes include identification of local inconsistencies of the photo response non-uniformity (PRNU) pattern [1, 2]; splicing detectors based on either JPEG features [3]; or steganography-inspired local descriptors [4, 5].

The final tampering map is typically obtained by comparing the real-valued *response map* of a forensic detector to a threshold. The distributions of the responses are difficult to calculate or estimate accurately [5], which limits the applicability of formal selection strategies like the well-established Neyman-Pearson test [6]. Ad-hoc post-processing is typically employed to eliminate excessive number of false positives.

The traditional threshold-based decision can be augmented with prior knowledge, e.g., using the Markovian prior [2, 7]. As a result, the decisions regarding individual authentication units are no longer independent, but consider also the detection scores of their neighbors. The problem resolves to minimizing an energy function that prefers larger solid areas and discourages isolated erroneous spots. Such an approach can also facilitate propagation of reliable detection scores into unreliable regions with dark, solid or saturated content [2].

This work was partly supported by NSFC (U1135001, 61332012) and Shenzhen R&D Program (GJHZ20140418191518323).

P. Korus and J. Huang are with the College of Information Engineering, Shenzhen University, Shenzhen, China and Shenzhen Key Laboratory of Media Security. P. Korus is also with the Department of Telecommunications, AGH University of Science and Technology, Kraków Poland, (e-mail: pkorus@agh.edu.pl,jwhuang@szu.edu.cn).

Supplementary materials, including the data set and a C++ implementation of the proposed method, can be obtained from http://kt.agh.edu.pl/~korus/.

Recently researchers began to explore new approaches that do not require explicit mapping of the detection scores to the final decisions. Chen and Hsu argue that the tampered region should constitute a sparse cluster of outliers in a sufficiently discriminative feature space [8]. Starting from robust principal component analysis, they derive an iterative algorithm that estimates the expected low-rank structure of pristine features.

In this paper, we propose to use the maximal entropy random walk (MERW) [9] on a graph to post-process the response map. Random walks, including MERW, have been successfully adopted for many problems including image segmentation [10], link prediction [11], and saliency estimation [12]. Our approach is inspired by recent successful use of MERW for detection of salient objects in digital photographs [13]. In conventional random walk, only local knowledge is available to the walker when making the transitions. In MERW the walker is aware of the full graph structure, which is beneficial for many applications. MERW is characterized by a strong localization property which will highlight important regions of the map, and attenuate the background.

Adjusting the weights of the graph, allows to obtain sensitivity to various aspects of the response map. To demonstrate the potential of this flexibility, we consider different weight computation strategies, including a visual saliency clue [13]. Distinct visibility of the tampered area in the response map has served as an *"informal visual proof"* of efficacy of newly proposed localization schemes. However, it has never been formally used for tampering localization - either separately or as side-information. We are aware of only one study that correlated saliency-based eye fixation coordinates with sharp edges indicative of crude splicing [14]. Flexibility of our approach allows us to formalize the above observation, and evaluate how a saliency-based model compares to conventional forensic clues. Sensitivity to different aspects of the response map makes the saliency clue a potentially interesting tool for decision fusion in the emerging multi-modal localization schemes. Dempster-Shafer Theory of Evidence identifies clue diversity as a critical property for decision fusion [15].

The proposed approach significantly outperforms existing methods, including the commonly used threshold-based decision with ad-hoc post-processing, and a recently proposed optimization-based approach with a Markovian prior [6, 7]. In our experimental evaluation on synthetic JPEG splicing forgeries, adoption of MERW yielded cleaner outputs and was particularly beneficial for noisy response maps.

The paper is organized as follows. Sections II and III introduce the MERW fundamentals, and the tampering localization protocol, respectively. The performed experimental evaluation is presented in Section IV. We conclude in Section V. Selected issues are extended in supplementary materials.

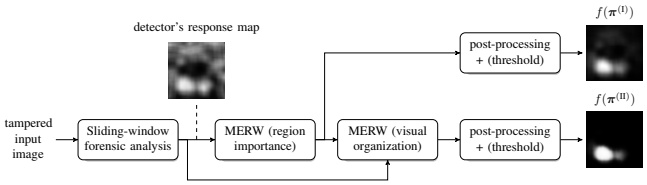

Fig. 1. Proposed localization framework; successive maximal entropy random walks highlight important regions and improve visual organization of the decision map; typical post-processing includes: map normalization, brightness correction, and optionally binarization, and morphological cleaning.

II. Maximal Entropy Random Walk

This section introduces the fundamental notions of MERW. For more information the readers can refer to [9, 16].

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote an undirected graph with vertices $\mathcal{V} = \{1, \ldots, N\}$ and edges $\mathcal{E} = \{(i,j) : i,j \in \mathcal{V}\}$. A random walk refers to a sequence of random variables $X_t \in \mathcal{V}$ where $t \in \{0, 1, \ldots\}$ indexes successive time instances. In every step, a walker performs a random transition from the current node $i \in \mathcal{V}$ to a neighboring node $j : j \in \mathcal{V} \wedge (i,j) \in \mathcal{E}$. The transitions are typically described by a probability matrix $\mathbf{P}_{N\times N}$ with elements $p_{ij} := P(X_{t+1} = j | X_t = i)$.

We consider a weighted graph, conveniently described by a weight matrix $\mathbf{W}_{N\times N}$. The weight of individual edges (i,j), denoted as w_{ij}, are non-negative and can be interpreted as the average number of available paths from node i to node j. A conventional random walker considers the nearest hop transitions to be equiprobable. A maximal entropy walker considers longer m-step paths passing through each prospective next node, which makes it aware of the global structure of the graph. The transition probabilities can be formally defined as:

$$p_{ij} = \frac{w_{ij} u_j}{\lambda u_i} \tag{1}$$

where λ is the largest eigenvalue of the weight matrix $\mathbf{W}$, and u_i, u_j denote the i-th and the j-th component of the corresponding eigenvector $\mathbf{u}$. Since the weight matrix is non-negative, both λ and $\mathbf{u}$ are non-negative according to the Frobenius-Perron theorem [9, 13]. In this study, the quantity of interest is the stationary probability density of finding the walker at every node in the graph $\pi_i : i \in \mathcal{V} \wedge \sum_i \pi_i = 1$. For MERW this probability is simply $\pi_i = u_i^2$ (generalizes to $\pi_i = u_i v_i$ for directed graphs with asymmetric weights having different left and reight eigenvectors $\mathbf{u}$ and $\mathbf{v}$ [16]).

MERW possesses a strong localization property, which will lead to suppression of low-degree ($\sum_j w_{ij}$) nodes and stronger highlighting of high-degree nodes. Hence, by properly choosing the weights of the graph, it is possible to exploit this phenomenon for exposition of important image regions.

III. MERW for Tampering Localization

The localization procedure is summarized in Fig. 1. The input to the model is a real-valued response map $\mathbf{c}$ of a forensic detector with detection scores for individual authentication units denoted as $c_i \in [0,1]$. The extremes of the response range represent the degree to which a certain forensic feature is present / absent (or alternatively the confidence of the classifier regarding the region's authenticity).

We span a fully-connected graph on non-overlapping image blocks of size 8×8 pixels (px)[1]. Similarly to [13], we consider a two-layer approach where two successive MERWs aim to identify important regions of the input map, and improve visual organization of the output map, respectively. We will separately assess the improvement of each of these layers. The stationary probability densities $\boldsymbol{\pi}^{(\mathrm{I})}$ and $\boldsymbol{\pi}^{(\mathrm{II})}$, after normalization to $[0,1]$ & prospective post-processing or binarization, will constitute the final decision map for the considered scenarios.

The first graph uses weights computed as:

$$w_{ij}^{(\mathrm{I})} = \phi(c_i, c_j) e^{-d_{ij}^2/\sigma_1^2} \ . \tag{2}$$

The scores c_i and c_j correspond to the source and the target nodes, respectively. The exponential term is responsible for fading of the weights with normalized L2 distance between the nodes $d_{ij} \in [0,1]$; σ_1^2 controls the fading strength. The main component of the weight $\phi(c_i, c_j)$ determines the sensitivity of the model to various aspects of the response map. The choice of this weight is discussed separately in Section III-A.

The second graph facilitates better visual organization of the tampering map, and helps to propagate the detected importance to similar regions nearby. It is particularly useful for saliency-related clues which tend to attenuate the middle of the detected objects [13]. This graph uses the following weights:

$$w_{ij}^{(\mathrm{II})} = (1 - |c_i - c_j|)\, \pi_i^{(\mathrm{I})} \pi_j^{(\mathrm{I})} e^{-d_{ij}^2/\sigma_2^2} \ , \tag{3}$$

where $\pi_i^{(\mathrm{I})}$ corresponds to the importance score of node i resulting from MERW on the first graph.

A. Selection of Graph's Weights

Complexity of the considered graph makes it impossible to compute the optimal weights analytically. Hence, we consider a few representative heuristic selection strategies motivated by the properties of MERW. We have also evaluated more sophisticated strategies. However, they provided marginal improvement and hence will not be presented in this paper.

Since the walker will concentrate on highly connected regions (localization property), greater weights should be assigned between tampered nodes. Edges between pristine regions should have low weights. Hence, one possible strategy is to make the weight propertional to the average response of the source and target node ($\propto c_i + c_j$). A more agggressive strategy would ignore the source node, and encourage transitions to target nodes with confident response ($\propto c_j$). It is also interesting to exploit visual saliency in the response maps, and adopt weights $\propto |c_j - c_i|$ (contrast between nodes). Hence, we finally consider:

- $\phi(c_i, c_j) = (c_i + c_j)/2$ - *mean response* mode;
- $\phi(c_i, c_j) = c_j$ - *raw target response* mode;
- $\phi(c_i, c_j) = |c_j - c_i|$ - *visual saliency* mode.

[1]Other possible choices include, e.g., superpixels [13]. While beneficial for shape identification and computational complexity, they are not necessarily applicable in forensics - for object insertion forgeries, certain improvement can be expected; for object removal, the effect is likely to be the opposite.

IV. Experimental Evaluation

This section describes the considered forensic detector, the adopted evaluation scenario, and the obtained results. Selected issues are extended in supplementary materials.

A. Detecting Forgeries in JPEG Images

We consider detection of JPEG image splicing forgeries with the use of mode-based first digit features [17]. We follow a similar localization scheme as Amerini et al. [3], but use overlapping analysis windows instead of non-overlapping ones. We consider 64×64 px blocks, moving with a single block stride (8 px). As a result, we obtain smooth response maps (column 2 in Fig. 3). We also increased the number of features to 180 (20 modes of 9 features each) and the density of the possible JPEG compression levels to $Q_1, Q_2 \in \mathcal{Q} = \{50, 51, \ldots, 100\}$. For each window size and each JPEG quality level Q_2, we train a separate support vector machine (SVM) to distinguish between single and double JPEG compression. The training was performed on 64×64 px windows randomly chosen from 1,338 images from the UCID dataset [18] and the relevant parameters of the SVM were found by a grid search with 3-fold cross-validation.

B. Experiment Set-up

We test the proposed method on a data set of 1,000 response maps obtained from synthetic splicing forgeries with the above detector. The images are of size 512×512 px and come from the BOSS data set [19]. Each forgery involves replacement of selected image regions (according to the patterns in col. 1 of Fig. 3) with the same content from the same image but with a different compression history. For each forgery the tampering pattern is randomly placed on the image plane. We include two scenarios with double-JPEG / single-JPEG compression artifacts inside / outside of the tampered regions. The first and second quality levels are chosen randomly from $\mathcal{Q}$ (but $Q_1 < Q_2$), leading to diverse response maps of varied reliability.

We compare our performance to traditional threshold-based decision, followed by removal of small connected components (having less than 16 image blocks), which along morphological erosion constitute the most popular approaches to cleaning the tampering maps of false positives. We also evaluate a more sophisticated method that includes the Markovian prior [2]. We resort to a popular Ising model with 1-st order neighborhood and minimize the corresponding energy function with a graph cuts-based solver [20, 21] from the UGM toolbox [22]. Localization trade-offs are controlled with three parameters: decision bias α, interaction strength β, and threshold τ. For this method, we refrain from ad-hoc map cleaning, and allow the Markovian prior to remove false-positives. To facilitate stronger neighborhood interactions, we post-process the resulting decision map with morphological dilation. We describe the adopted model in detail in supplementary materials.

To observe localization performance trade-offs, we sweep the decision threshold τ, and measure the number of true positives (t_p), false negatives (f_n), false positives (f_p) and the $F_1 = 2\ t_p(2\ t_p + f_n + f_p)^{-1}$ score. The threshold changes non-uniformly in the range $[0, 1]$ to increase the number of samples with low false positive rates. The parameters for the localization methods were found by grid search on a subset of 40 diverse response maps (Tab. I). Details of the parameter search are included in supplementary materials.

TABLE I
Parameter selection results for the energy minimization approach $(\hat{\alpha}, \hat{\beta})$ and MERW with various weights (σ_1^2, σ_2^2).

Method / weight mode	Best parameters	Grid range	Step
Markov	$-2.5,\ 2.7$	$[-3, 3] \times [0, 3]$	linear
Optimization for the best AUC after the first MERW			
saliency	0.37	$[10^{-2}, 1]$	log
target response	0.10	$[10^{-2.5}, 1]$	
mean response	0.06		
Optimization for the best AUC after the second MERW			
saliency	$0.67,\ 0.033$	$[10^{-2}, 1]^2$	log
target response	$1.00,\ 0.023$	$[10^{-2.5}, 1]^2$	
mean response	$0.29,\ 0.030$		

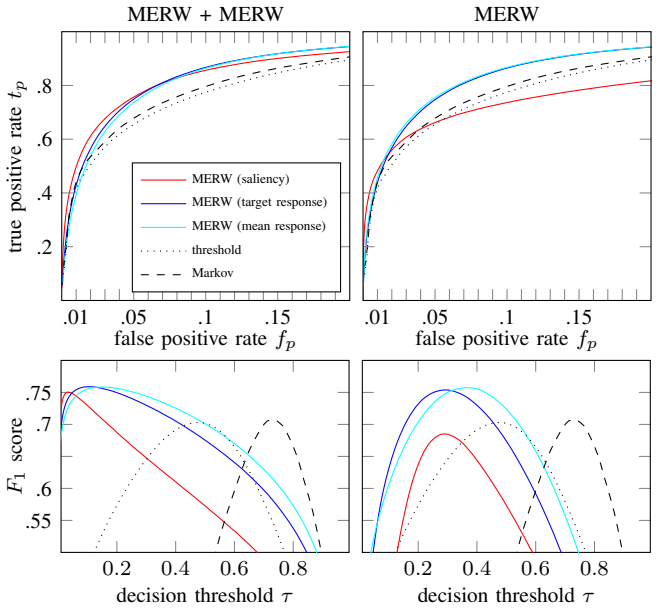

Fig. 2. Localization performance for 1,000 synthetic JPEG forgeries: receiver operation characteristics (top); F_1 score curves (bottom).

C. Results

Fig. 2 shows the obtained receiver operation characteristics (ROC) and the $F_1(\tau)$ curves for the considered localization schemes. For our tampering scenario, incorporation of the Markovian prior turned out to provide little benefits over the conventional threshold with ad-hoc post-processing ($F_1 = 0.703$ vs. 0.707). We attribute this phenomenon to three main reasons. Firstly, high variety of the response maps and tampering patterns makes it difficult to choose a single universal set of parameters. Along with false positives the Markovian prior removed fine details of the tampered regions' shapes. Secondly, high overlap of the sliding window introduced spatial correlations to the response maps. Hence, even the traditional threshold-based decision already incorporated some

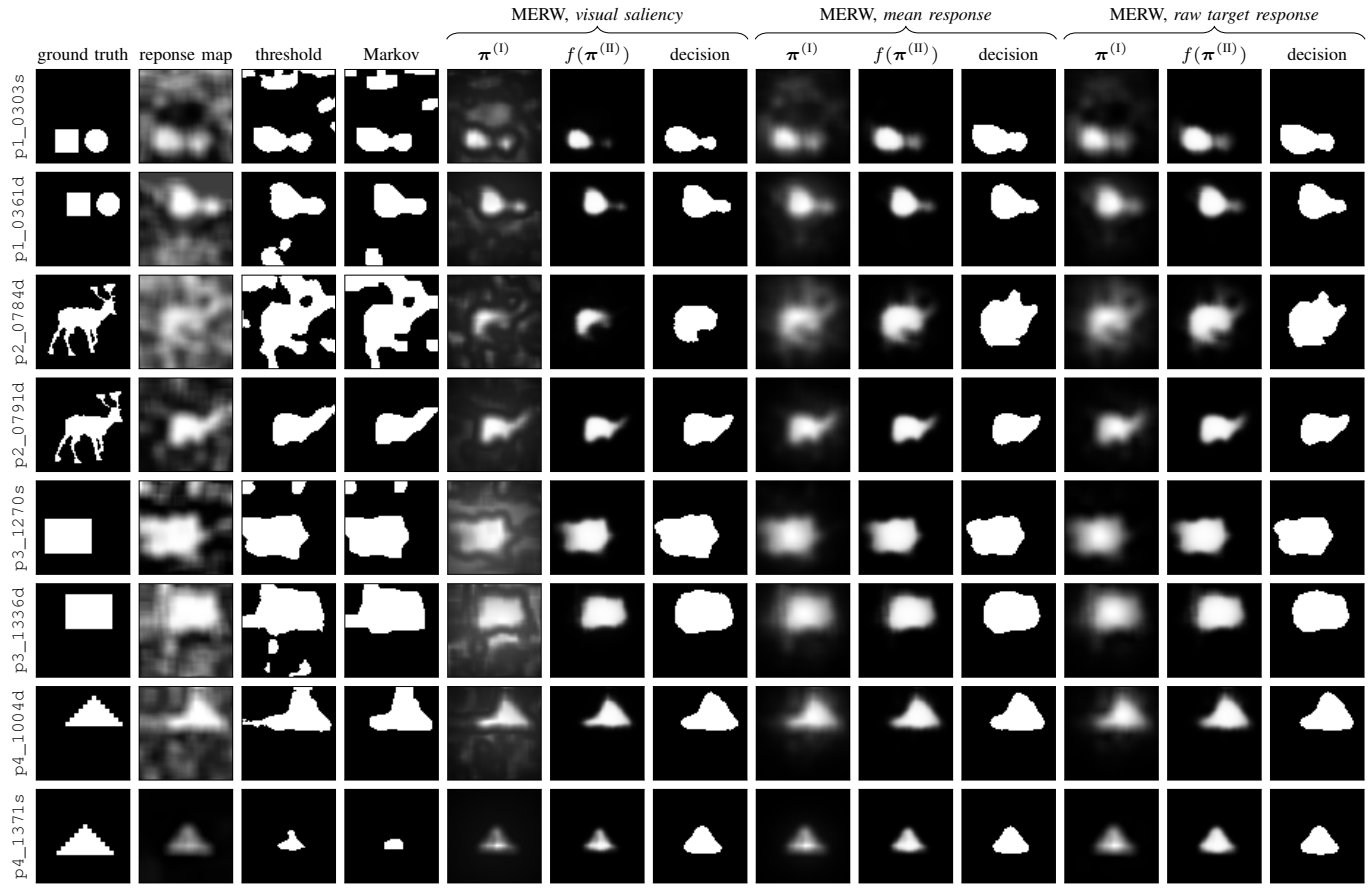

Fig. 3. Example tampering localization results (labeled vertically with file names from the data set); the first columns illustrates the considered tampering patterns (*p1 - p4*); the post-processing function was $f(x) = 1-e^{-5x}$; binary decision thresholds were chosen to match the best average F_1 scores: $\tau = 0.473$ (threshold); $\tau = 0.722$ (Markov); $\tau = 0.035$ (MERW, *visual saliency*); $\tau = 0.10$ (MERW, *mean response*); $\tau = 0.15$ (MERW, *target response*).

information about the forensic scores in the neighborhood. Finally, our forensic detector already produced solid shapes without holes that the Markovian prior could fill.

Adoption of MERW is clearly beneficial - it considerably improves the localization performance both in terms of ROCs and F_1 scores. The impact of the weight selection strategies remains limited. Both the *mean* and *raw target response* strategies yielded comparable results, with minor advantage of the latter. Incorporation of the second MERW brings little improvement in terms of ROC performance, but significantly impacts the appearance of the map. The resulting maps are more attenuated, which can be observed both in the example tampering maps (Fig. 3) and in the shifted maximum of the $F_1(\tau)$ curve towards $\tau \to 0$. Non-linear brightness adjustment might be desirable for the convenience of subsequent processing and display. We used a simple mapping $f(x) = 1-e^{-5x}$ to boost low scores in the final map. Adjusting the strength of this mapping does not change the ROC performance, and allows to obtain a clearer map with flatter F_1 curve. Incorporation of the visual saliency clue reveals different behavior. It outperformed other strategies for low false positive rates, but required the second MERW for adequate results.

Fig. 3 shows example tampering maps for the considered localization schemes. It shows the real-valued maps produced by both MERWs (with separately chosen parameters for the single / double MERW) and the final decision obtained from the second one. Notice how weight selection impacts importance of different map regions. In particular, note how the saliency model responds to distinctive, but not necessarily white, regions of the response maps (e.g., just above the tampered area for `p1_0303s`). In general, rationale for such behavior depends on the forensic feature at hand. It might be desirable to attract the attention of either a human analyst or successive processing steps to response map anomalies, and then to cross reference these regions with other image features.

V. Conclusions

Our study shows that the maximal entropy random walk can be successfully adopted for tampering localization in digital image forensics. For synthetic JPEG splicing forgeries, it significantly outperformed existing localization strategies, including the commonly used threshold-based decision, and the recently proposed energy minimization approach with a Markovian prior. The proposed method servers as an additional post-processing step, and does not require any changes to existing forensic detectors. Adjusting the weights of the constructed graph allows to use various clues from the input response map. To demonstrate flexibility of the proposed approach, we evaluated three distinct weight selection strategies - each of which performed better than the state-of-the-art.

REFERENCES

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

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

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

[4] D. Cozzolino, D. Gragnaniello, and L. Verdoliva, "Image forgery detection based on the fusion of machine learning and blockmatching methods," http://arxiv.org/abs/1311.6934, 2013.

[5] L. Verdoliva, D. Cozzolino, and G. Poggi, "A feature-based approach for image tampering detection and localization," in *Proc. of IEEE Int. Workshop on Information Forensics and Security*, 2014.

[6] G. Chierchia, D. Cozzolino, G. Poggi, C. Sansone, and L. Verdoliva, "Guided filtering for PRNU-based localization of small-size image forgeries," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2014, pp. 6231–6235.

[7] W. Wang, J. Dong, and T. Tan, "Exploring DCT coefficient quantization effects for local tampering detection," *IEEE Trans. on Information Forensics and Security*, vol. 9, no. 10, pp. 1653–1666, 2014.

[8] Y.-L. Chen and C.-T. Hsu, "What has been tampered? from a sparse manipulation perspective," in *Proc. of IEEE Int. Workshop on Multimedia Signal Processing*, Sept 2013, pp. 123–128.

[9] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, "Localization of maximal entropy random walk," http://arxiv.org/abs/0810.4113, 2008.

[10] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen, "Random walk and graph cut for co-segmentation of lung tumor on pet-ct images," *Image Processing, IEEE Transactions on*, vol. 24, no. 12, pp. 5854–5867, Dec 2015.

[11] R.-H. Li, J. X. Yu, and J. Liu, "Link prediction: the power of maximal entropy random walk," in *Proc. of the ACM int. conf. on Information and knowledge management*. ACM, 2011, pp. 1147–1156.

[12] V. Gopalakrishnan, Yiqun Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *Image Processing, IEEE Transactions on*, vol. 19, no. 12, pp. 3232–3242, Dec 2010.

[13] J.-G. Yu, J. Zhao, J. Tian, and Y. Tan, "Maximal entropy random walk for region-based visual saliency," *IEEE Tran. on Cybernetics*, vol. 44, no. 9, pp. 1661–1672, 2014.

[14] Z. Qu, G. Qiu, and J. Huang, "Detect digital image splicing with visual cues," in *Information Hiding*, vol. 5806 of *Lecture Notes in Computer Science*, pp. 247–261. 2009.

[15] 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 Trans. on Information Forensics and Security*, vol. 8, no. 4, pp. 593–607, 2013.

[16] J. Duda, *Extended Maximal Entropy Random Walk*, Ph.D. thesis, Jagiellonian University, 2012.

[17] B. Li, Y.Q. Shi, and J. Huang, "Detecting doubly compressed jpeg images by using mode based first digit features," in *Proc. IEEE Workshop on Multimedia Signal Processing*, Oct 2008, pp. 730–735.

[18] G. Schaefer and M. Stich, "Ucid - an uncompressed colour image database," in *Proc. SPIE Storage and Retrieval Methods and Applications for Multimedia*, 2004, pp. 472–480.

[19] "The dataset from the break our steganographic system contest," http://www.agents.cz/boss/index.php, 2010, Visited on 26 March 2015.

[20] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov 2001.

[21] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.

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