Rotation Invariant Localization of Duplicated Image Regions Based on Zernike Moments

Seung-Jin Ryu, Matthias Kirchner, Min-Jeong Lee, and Heung-Kyu Lee

***Abstract*—This paper proposes a forensic technique to localize duplicated image regions based on Zernike moments of small image blocks. We exploit rotation invariance properties to reliably unveil duplicated regions after arbitrary rotations. We devise a novel block matching procedure based on locality sensitive hashing and reduce false positives by examining the moments' phase. A massive experimental test setup benchmarks our algorithm against state-of-the-art methods under various perspectives, examining both pixel-level localization and image-level detection performance. By taking signal characteristics into account and distinguishing between "textured" and "smooth" duplicated regions, we find that the proposed method outperforms prior art in particular when duplicated regions are smooth. Experiments indicate high robustness against JPEG compression, blurring, additive white Gaussian noise, and moderate scaling.**

***Index Terms*—Copy-move detection, duplicated region localization, locality sensitive hashing, Zernike moments.**

I. Introduction

EASY-TO-USE imaging devices and inexpensive storage space make the acquisition of high-quality digital images a natural form of human perception of and interaction with the real world. At the same time, the very nature of digital data puts into question many of the positive aspects that we usually associate with digital images. Digital data can be manipulated easily. This raises questions regarding the authenticity of digital images, and a constantly growing number of uncovered manipulations [1] is certainly only the tip of the iceberg.

Within the last decade, scholars in digital image forensics have set out to develop *passive-blind* techniques to restore some of the lost trustworthiness of digital images [2]. These methods work by examining inherent statistical characteristics of an image in question. One of the key assumptions of passive—blind forensics is that processed images exhibit *artifacts*, e.g., due to resampling [3], [4], double quantization [5], [6], contrast enhancement [7], smoothing [8], or sharpening [9]. The presence of such artifacts in a questioned image may indicate a manipulation. While a strict line between legitimate and illegitimate processing is often hard to draw in practice [10], *local* processing artifacts are often deemed particularly critical.

Copy-move (CM) manipulations are a common form of local processing, where parts of an image are copied and reinserted into another part of the same image [11], often to conceal or emphasize image details. While basic CM manipulations are typically considered to duplicate parts without further change, realistic manipulations may also require a geometric transformation of the original part prior to reinsertion. This can result in a better alignment of the duplicated part with its surrounding. In the following, we focus on *rotation* as a typical representative of such transformations. We refer to a copy—move manipulation that involves rotated duplicate regions as *copy-rotate-move* (CRM) manipulation.

A number of passive—blind methods to unveil copy—(rotate)—move manipulations have already been proposed. Most of them adopt a procedure first outlined in Fridrich *et al.*'s [11] seminal CM paper. Section II recalls the general scheme before we follow its trail to present a CRM detector based on Zernike moments [12] of small overlapping image blocks. These moments have already found wide application in areas such as pattern recognition and digital watermarking. Here, we demonstrate their superior performance as building blocks of CRM detectors. An overview of relevant properties of Zernike moments in Section III lays the foundation for their application to CRM detection in Section IV. Section V benchmarks our detector with state-of-the-art methods and presents empirical evidence from a massive test setup. Section VI concludes the paper.

II. Detecting Copy(—Rotate)—Move Manipulations

A. General Detection Procedure

Copy—move manipulations result in (near-)duplicate image regions, which practical forensic analyses examine in terms of *robust feature representations* of parts of the image. Fig. 1 details the principal detection pipeline [13]. After optional preprocessing (e.g., color to grayscale conversion), the image is transformed to the feature space. A set of *feature vectors* represents *local image characteristics* and is inspected for similarities in a *matching procedure*. This is achieved either by splitting the image into small blocks, which are then transformed separately, or by finding salient key points and extracting feature vectors based thereon. The detector then outputs tuples of similar feature vectors or their corresponding coordinates in the image plane. False positives in the matching procedure are pruned in a final *error reduction* step.

Manuscript received December 26, 2012; revised April 01, 2013 and July 01, 2013; accepted July 01, 2013. Date of publication July 04, 2013; date of current version July 12, 2013. This work was supported in part by Defense Acquisition Program Administration and Agency for Defense Development under Contract (UD060048AD), and in part by the Korean government (MEST) Grant 2012R1A2A1A05026327. The work of M. Kirchner was supported by a postdoctoral fellowship at the International Computer Science Institute in Berkeley, CA, USA, funded by the German Academic Exchange Service (DAAD). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jiwu Huang. *(Corresponding author: H.-K. Lee.)*

S.-J. Ryu and H.-K. Lee are with the Department of Computer Science, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (e-mail: sjryu@mmc.kaist.ac.kr; hklee@mmc.kaist.ac.kr).

M. Kirchner is with the Department of Information Systems, WWU Münster, 48149 Münster, Germany (e-mail: matthias.kirchner@uni-muenster.de).

M.-J. Lee is with the Information Technology R&D Center, SK Telecom, Seoul, 100-999, Korea (e-mail: mjlee@mmc.kaist.ac.kr).

Digital Object Identifier 10.1109/TIFS.2013.2272377

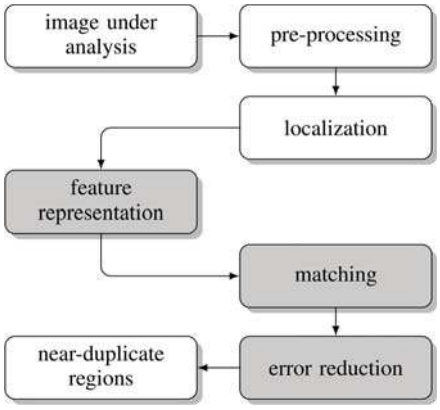

Fig. 1. General copy(—rotate)—move detection pipeline [13]. Highlighted blocks refer to components where this manuscript presents advances.

B. Copy—Move Manipulation

Specific CM detection features include quantized DCT coefficients of small blocks [11], [14], or the representation of relevant block characteristics by means of a principal component analysis (PCA) or kernel PCA [15], [16]. Other approaches measure the similarity of blocks based on Hu moments [17], discrete wavelet transform [16], singular value decomposition [18], intensity-based features [19], or blur-invariant moments [20]. As to the matching procedure, most techniques lexicographically sort the respective feature vectors and measure the similarity between adjacent sorted vectors [11]. However, lexicographical sorting is suboptimal in this respect as adjacent sorted vectors are not necessarily near-duplicate. kd-trees give better performance [21], yet at the cost of increased computational complexity. Error reduction typically involves the search for pairs of feature vectors that are linked via the same translation vector in the image plane. Assuming that the duplicated region covers a reasonable number of blocks or key points, a peak in the histogram of translation vectors hints to the spatial relationship of original and duplicated region.

C. Copy—Rotate—Move Manipulation

The procedure in Fig. 1 also applies to most CRM detectors. To account for rotated regions, feature representations have to be chosen invariant to rotation. Bayram *et al.* [22] use the Fourier-Mellin transformation, yet experiments indicate reliability only for small rotation angles. Li *et al.* [23] improve Bayram's method, but the reported detection accuracy still remains rather low. In a similar attempt, Bravo-Solorio *et al.* [24] represent blocks in log-polar coordinates and construct a feature vector by angular integration over fixed radii. However, as this representation, as well as a recent extension by Wu *et al.* [25], directly depends on individual pixel intensities, it is particularly sensitive to changes thereof. Wang *et al.* [26] propose a more robust approach based on Gaussian pyramids and mean intensities of circular blocks, which Liu *et al.* [27] combine with Hu moments. An alternative stream of research builds upon scale-invariant feature transforms (SIFT) and feature vectors from the neighborhood of geometrically invariant key points [28]–[30]. Yet such schemes have in common that key points can only be computed from salient image regions, which can pose a considerable drawback in practical analyses.

CRM manipulations further require a generalized procedure to identify matched blocks or key points across duplicated regions. A common approach is to estimate the underlying affine transformation by fitting a system of linear equations to the corresponding coordinates in the image plane. Pairs of feature vectors are considered to be part of a duplicated region if the number of pairs with equal transformation parameters exceeds a certain threshold [13], and/or estimated transformation parameters do not deviate substantially [29].

D. Main Contributions

Within this paper, four major contributions greatly advance the field of C(R)M detection and overcome some of the limitations discussed above. (See also Fig. 1, where the corresponding components in the detection pipeline are printed on gray background.) First, we represent individual blocks by Zernike moments [12] up to an appropriate order. A rotation-invariant magnitude makes these moments particularly promising *CRM detection features* [31]. The second contribution is an *efficient block matching procedure* based on locality sensitive hashing (LSH) [32]. Third, we incorporate the phase of Zernike moments into a *feature space error-reduction* procedure. This differs from earlier spatial domain approaches and yields increased accuracy. Finally, a massive test setup based on a set of 1000 images results in a *comprehensive reference benchmark* of state-of-the-art CRM detection methods. It is, to the best of our knowledge, the first of this kind in the literature. In particular, we explicitly take signal characteristics into account and distinguish between 'smooth' and 'textured' duplicated regions. Experiments consider both pixel-level localization and image-level detection of manipulations, respectively, different sizes of images and duplicated regions, as well as robustness against a number of representative distortions.

III. Zernike Moments

Moments and invariant functions of moments have been used extensively for feature extraction in a wide range of pattern recognition and digital watermarking applications [33], [34]. Among the various types of moments discussed in the literature, in particular Zernike moments [12] are known for their superior insensitivity to image noise, their information content, and their ability to provide robust image representations [34]–[36]. This section recalls the mathematical foundations of Zernike moments based on [34], [36].

A. Definition

Let $0 \leq \rho \leq 1$ and $0 \leq \theta \leq 2\pi$ represent polar coordinates over the closed unit disk. For a continuous image function $f(\rho, \theta)$ that vanishes outside the unit disk, the Zernike moment of order n with repetition m is defined as

$$A_{n,m} = c_n \int_0^{2\pi} \int_0^1 f(\rho, \theta) V^*_{n,m}(\rho, \theta) \rho \mathrm{d}\rho \mathrm{d}\theta. \tag{1}$$

Here, n is a nonnegative integer, and m is an integer such that $n - |m|$ is nonnegative and even. Coefficient c_n is given by

$$c_n = \frac{n+1}{\pi}. \tag{2}$$

Computing Zernike moments requires evaluating complex functions $V_{n,m}(\rho,\theta)$, defined as

$$V_{n,m}(\rho,\theta) = R_{n,m}(\rho)e^{jm\theta}, \tag{3}$$

with Zernike polynomials [37] $R_{n,m}(\rho)$ given by

$$R_{n,m}(\rho) = \sum_{s=0}^{\frac{(n-|m|)}{2}} \frac{(-1)^s\,[(n-s)!]\,\rho^{n-2s}}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}. \tag{4}$$

These polynomials are orthogonal and satisfy

$$\int_0^{2\pi}\int_0^1 V_{n,m}^*(\rho,\theta)\times V_{p,q}^*(\rho,\theta)\mathrm{d}\rho\mathrm{d}\theta = \frac{\delta_{n,p}\delta_{m,q}}{c_n}, \tag{5}$$

where $\delta_{a,b}$ is the Kronecker delta:

$$\delta_{a,b} = \begin{cases} 1 & \text{for } a=b, \\ 0 & \text{else.} \end{cases} \tag{6}$$

It follows from (4) that $R_{n,-m}(\rho) = R_{n,m}(\rho)$ and thus $A_{n,m}^* = A_{n,-m}^*$.

B. Rotation Invariance of Zernike Moments

A desirable property of Zernike moments is their analytical invariance to rotation. Consider rotation of an image f by angle α. Writing the rotated image as f', the relationship between original and rotated image is

$$f'(\rho,\theta) = f(\rho,\theta-\alpha). \tag{7}$$

Combining (1) and (3) to

$$A_{n,m} = c_n\int_0^{2\pi}\int_0^1 f(\rho,\theta)R_{n,m}(\rho)\rho e^{-jm\theta}\mathrm{d}\rho\mathrm{d}\theta, \tag{8}$$

the Zernike moments of the rotated image are given by

$$A'_{n,m} = c_n\int_0^{2\pi}\int_0^1 f(\rho,\theta-\alpha)R_{n,m}(\rho)\rho e^{-jm\theta}\mathrm{d}\rho\mathrm{d}\theta. \tag{9}$$

Substituting $\theta_1 = \theta-\alpha$, we obtain

$$A'_{n,m} = c_n\int_0^{2\pi}\int_0^1 f(\rho,\theta_1)R_{n,m}(\rho)\rho e^{-jm(\theta_1+\alpha)}d\rho d\theta_1 \tag{10}$$

$$= c_n e^{(-jm\alpha)}\int_0^{2\pi}\int_0^1 f(\rho,\theta_1)R_{n,m}(\rho)\rho e^{-jm\theta_1}d\rho d\theta_1 \tag{11}$$

$$= A_{n,m}e^{-jm\alpha}. \tag{12}$$

Hence, rotation of an image translates to a phase shift in the corresponding Zernike moments. It follows that the magnitude of Zernike moments, $|A_{n,m}|$, can be interpreted as rotationally invariant feature representation of a given image.

C. Rotation Angle Estimation Using Zernike Moments

It is possible to determine the rotation angle between two images by analyzing the phase of their Zernike moments [38]. By making use of (12), the Zernike moments of a rotated image can be written as

$$A'_{n,m} = A_{n,m}e^{-jm\alpha} \tag{13}$$

$$= |A_{n,m}|e^{j\phi_{A_{n,m}}}e^{-jm\alpha} \tag{14}$$

$$= |A_{n,m}|e^{j\phi_{A'_{n,m}}}. \tag{15}$$

The phase of these moments can be expressed as

$$\phi_{A'_{n,m}} = \phi_{A_{n,m}} - m\alpha, \tag{16}$$

so that the phase difference, $\Phi_{n,m}$, is given by

$$\Phi_{n,m} \equiv \left(\phi_{A_{n,m}} - \phi_{A'_{n,m}}\right) \bmod 2\pi \tag{17}$$

$$\equiv m\alpha \bmod 2\pi. \tag{18}$$

Because $\forall n\in\mathbb{Z}: \Phi_{n,1}\equiv\alpha \bmod 2\pi$, the rotation angle α can be inferred from (18) via examination of arbitrary n-th order moments with repetition $m=1$.

IV. CRM Detection based on Zernike Moments

Analytical invariance to rotation and robustness to noise make Zernike moments a perfect building block for detecting copy—rotate—move (CRM) manipulations. This section details a specific instance of such detectors. We extract Zernike moments from overlapping blocks of a questioned image and use their magnitudes as feature representation. The detector employs locality sensitive hashing (LSH) [32] for block matching and removes falsely matched block pairs by inspecting phase differences of corresponding Zernike moments.

A. Feature Extraction

Given an $M\times N$ grayscale image under analysis, we denote overlapping blocks of size $L\times L$ as $f^{(i,j)}(x,y)$, where superscript $(i,j)\in\{0,\ldots,M-L\}\times\{0,\ldots,N-L\}$ refers to row and column index of a block's upper-left corner in the intensity plane, respectively. Block size L is chosen under the assumption that the duplicated region spans more than a single block. Each of the N_{blocks} blocks in the image,

$$N_{\text{blocks}} = (M-L+1)\cdot(N-L+1), \tag{19}$$

is transformed to the feature space by computing its Zernike moments $A_{n,m}^{(i,j)}$. More specifically, we use (1)—with integrals replaced by summations—to obtain vectors $\boldsymbol{z}^{(i,j)}$ with the magnitudes of Zernike moments up to particular order $\bar{n}$,

$$\boldsymbol{z}^{(i,j)} = \left(\left|A_{0,0}^{(i,j)}\right|, \left|A_{1,1}^{(i,j)}\right|, \ldots, \left|A_{\bar{n},\bar{n}-2}^{(i,j)}\right|, \left|A_{\bar{n},\bar{n}}^{(i,j)}\right|\right)^\top. \tag{20}$$

Each vector $\boldsymbol{z}^{(i,j)}$ holds a total of N_{moments} elements,

$$N_{\text{moments}} = \sum_{k=0}^{\bar{n}}\left(\left\lfloor\frac{k}{2}\right\rfloor+1\right). \tag{21}$$

Block-based CRM detection then boils down to the examination of a set $\mathbb{Z}$ of N_{blocks} feature vectors for near-duplicate entries,

$$\mathbb{Z} = \{\boldsymbol{z}^{(0,0)}, \ldots, \boldsymbol{z}^{(M-L+1,N-L+1)}\}. \tag{22}$$

Parameter $\bar{n}$ trades off the robustness and the degree of detail captured by the feature representation. By the orthogonality property of (5), a block $f^{(i,j)}(x,y)$ can be approximated from its Zernike moments as

$$\hat{f}^{(i,j)}(x,y) = \sum_{n=0}^{\bar{n}} \sum_{m} A_{m,n}^{(i,j)} V_{n,m}^{(i,j)}(\rho,\theta). \tag{23}$$

The larger the maximum order $\bar{n}$ is, the more high-frequency information is taken into account [36]. Yet with $\bar{n}$ becoming too large, also the sensitivity to noise increases.

For practical computation of Zernike moments we take the block center as origin, mapping pixel coordinates $(x,y) \in \{0,\ldots,L-1\}^2$ to the range of the unit circle. Our implementation ignores pixels outside the unit circle. This common procedure introduces a geometric approximation error. It occurs in combination with numerical discretization errors and is relatively stronger for smaller block sizes and higher-order moments [39].

B. Locality Sensitive Hashing

Because CRM manipulations are likely to result in pairs of similar feature vectors, we inspect set $\mathbb{Z}$ for near-duplicate entries $(\boldsymbol{z}^{(i,j)}, \boldsymbol{z}^{(k,l)})$ within the Euclidean distance

$$\|\boldsymbol{z}^{(i,j)} - \boldsymbol{z}^{(k,l)}\| \leq D_1 \tag{24}$$

in the feature space. We adopt locality sensitive hashing (LSH) [32] to find matching blocks (i.e., pairings that adhere to (24)) with *success probability* $\sigma \geq 1-\varepsilon$ for a given error probability $\varepsilon > 0$. This works by mapping each entry of set $\mathbb{Z}$ to a (sequence of) hash table(s), which can then be searched for candidate blocks efficiently. A hash function $h(\boldsymbol{z})$ is called locality sensitive, if the probability that two input vectors collide strictly decreases as a function of their distance. We exploit that such hash functions can be constructed from quantized random projections with quantization bin size w [40],

$$h^{(\boldsymbol{a},b)}(\boldsymbol{z}^{(i,j)}) = \left\lfloor \frac{\boldsymbol{a} \cdot \boldsymbol{z}^{(i,j)} + b}{w} \right\rfloor, \tag{25}$$

if the elements of random vector $\boldsymbol{a} = (a_0, \ldots, a_{N_{\text{moments}}-1})$ are independently drawn from an s-stable distribution. We use i.i.d. Gaussian $a_k \sim \mathcal{N}(0,1)$ (i.e., stability parameter $s=2$), with offset b drawn from a uniform distribution, $b \sim \mathcal{U}(0,w)$. Under this construction, the collision probability γ_h for two vectors $(\boldsymbol{z}^{(i,j)}, \boldsymbol{z}^{(k,l)})$ with Euclidian distance d is given by [40]

$$\begin{aligned}\gamma_h &= \gamma_h(d) = \Pr\left(h\left(\boldsymbol{z}^{(i,j)}\right) = h\left(\boldsymbol{z}^{(k,l)}\right)\right) \\ &= \frac{1}{d}\int_0^w (1 - tw^{-1}) p(td^{-1}) \mathrm{d}t,\end{aligned} \tag{26}$$

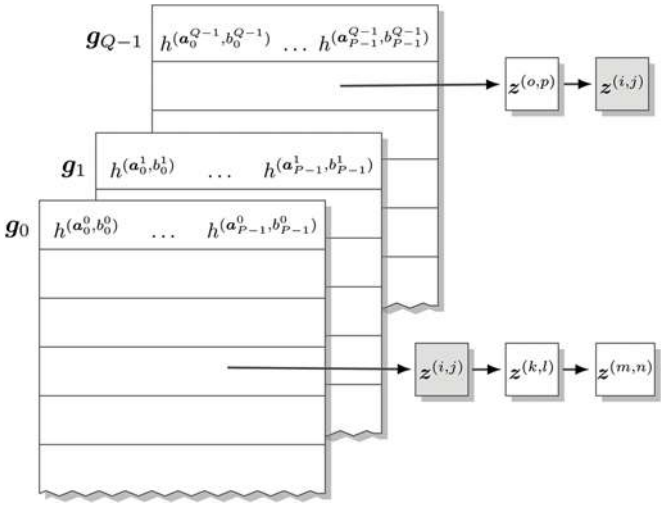

Fig. 2. Structure of hash tables for locality sensitive hashing.

where $p(x) = \sqrt{(2/\pi)}\exp(-x^2/2)$ denotes the density function of the half-normal distribution with standard deviation 1.

The collision probability for blocks with large feature space distance is reduced through P independent evaluations of (25), $\boldsymbol{g}(\boldsymbol{z}) = (h^{(\boldsymbol{a}_0,b_0)}(\boldsymbol{z}), \ldots, h_{P-1}^{(\boldsymbol{a},b_{P-1})}(\boldsymbol{z}))$. This procedure is repeated Q times to further increase the clustering accuracy for near-duplicate vectors. Hence, overall Q hash tables are constructed, whereas each feature vector $\boldsymbol{z}$ is stored in corresponding buckets $\boldsymbol{g}_0(\boldsymbol{z}), \ldots, \boldsymbol{g}_{Q-1}(\boldsymbol{z})$ [40]. Fig. 2 depicts the structure of hash tables.

As a result, the probability that two near-duplicate vectors do not collide in *any* of the Q tables is at most $\left(1-\gamma_h(D_1)^P\right)^Q$. Requiring a success probability $\sigma \geq 1-\varepsilon$, i.e.,

$$1 - \left(1 - \gamma_h(D_1)^P\right)^Q \geq 1 - \varepsilon, \tag{27}$$

establishes the following relation between the parameters of the hashing scheme:

$$Q \geq \frac{\log \varepsilon}{\log\left(1 - \gamma_h(D_1)^P\right)} = Q_{\min}. \tag{28}$$

Because the amount of time required to set up and to search the hash tables increases with Q, we set $Q = \lceil Q_{\min} \rceil$ [40].

Once the hash tables are initialized, finding a near-duplicate block for a query $\boldsymbol{z}^{(\text{query})} \equiv \boldsymbol{z}^{(i,j)}$ means inspecting all buckets $\boldsymbol{g}_0(\boldsymbol{z}^{(\text{query})}), \ldots, \boldsymbol{g}_{Q-1}(\boldsymbol{z}^{(\text{query})})$ for a feature vector $\boldsymbol{z}^{(\text{match})} \equiv \boldsymbol{z}^{(k,l)}$. As blocks in close spatial proximity are likely to yield relatively similar Zernike moments, we further evaluate the spatial distance between blocks in the intensity plane and require

$$\|(i,j) - (k,l)\| \geq D_2. \tag{29}$$

Among candidate blocks satisfying (29), the pair with minimum distance in the feature space is selected and considered as potentially being part of a duplicated region.

In practice, the parameters of this matching procedure are chosen commensurate with the forensic investigator's requirements on reliability and robustness. A small feature space dis-

tance D_1 lowers the number of erroneous matches, whereas larger values facilitate more robust decisions. Also a large spatial distance D_2 may reduce false matches, however at the cost of an increased chance of missing duplicated blocks nearby. For a desired success probability σ, the length P of hash vectors $\boldsymbol{g}(\boldsymbol{z})$ is central to the collision probability of the hashing scheme. It also trades off the time required for hash table construction and search, respectively. Shorter vectors mean faster construction, whereas we prefer rather large P to reduce search time over all blocks. We refer to Section V-A for the numerical parametrization used in our experiments.

C. Postprocessing to Reduce False Matching

False matching occurs when pairs of original blocks have similar magnitudes of Zernike moments despite not being duplicated. This section describes a procedure to remove such false positives from the set of potential CRM blocks. We inspect the set of matched block pairs for a common rotation angle by customizing the Random Sample Consensus (RANSAC) [41] algorithm with Zernike phase differences. Note that this feature space approach differs from prior art, where RANSAC has been applied directly to the corresponding block coordinates in the intensity plane [29], [30].

Algorithm 1 details our error reduction method. Denoting the set of all LSH-matched block pairs as $\mathbb{P}$, we estimate the rotation angle α_{Z} for all pairs in $\mathbb{P}$ by making use of (18),

$$\Phi_{1,1} = \alpha_{\mathrm{Z}} \bmod 2\pi. \tag{30}$$

Iterating over a set of candidate angles, we then conduct the following steps. Identifying block pairs for which the estimated rotation angle satisfies

$$|\alpha_{\mathrm{Z}} - \tilde{\alpha}| < T_1 \tag{31}$$

for a chosen angle $\tilde{\alpha}$ and threshold T_1 results in a set of block pairs $\mathbb{P}_{\tilde{\alpha}}$. A reduced set $\mathbb{P}^*_{\tilde{\alpha}}$ is constructed from $\mathbb{P}_{\tilde{\alpha}}$ by keeping only those pairings for which at least one spatially adjacent block pair is also included in $\mathbb{P}_{\tilde{\alpha}}$. Assuming that the block size is smaller than the duplicated region, the rationale is that spatial neighbors of a duplicated block are with high probability part of the same duplicated region. Finally, RANSAC is applied to the reduced set $\mathbb{P}^*_{\tilde{\alpha}}$. Denoting a pair of matched blocks as $(\boldsymbol{z}^{(i,j)}, \boldsymbol{z}^{(k,l)})$, we select three spatially adjacent colinear pairs from $\mathbb{P}^*_{\tilde{\alpha}}$ to infer their 2×2 affine transformation $\boldsymbol{M}$ in the spatial domain,

$$(i,j)^\top = \boldsymbol{M} \cdot (k,l)^\top + \boldsymbol{t}, \tag{32}$$

with $\boldsymbol{t}$ being a 2×1 translation vector. The polar decomposition $\boldsymbol{M} = \boldsymbol{S}\boldsymbol{R}$ is then used to extract the rotational part of matrix $\boldsymbol{M}$ [42]. In analogy to the polar form of a nonzero complex number, this factorizes $\boldsymbol{M}$ into a positive semidefinite "stretching" matrix $\boldsymbol{S}$ and an orthogonal "rotation" matrix $\boldsymbol{R}$.[1] The decomposition can be obtained in terms of the singular value decomposition (SVD) $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$,

$$\boldsymbol{M} = \boldsymbol{S}\boldsymbol{R} = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top)(\boldsymbol{U}\boldsymbol{V}^\top). \tag{33}$$

[1] $\boldsymbol{S}$ is not necessarily only a scaling matrix as it can have nonzero elements along the off-diagonal. Also, $\boldsymbol{R}$ may contain a reflection component.

For block pairs belonging to a duplicated region, we expect both that $\boldsymbol{S}$ is a scaling matrix and that rotation angle

$$\alpha_{\mathrm{R}} = \arccos R_{11} \tag{34}$$

conforms to candidate angle $\tilde{\alpha}$. Hence, RANSAC proceeds iff

$$\left\|\boldsymbol{I} - S_{11}^{-1}\boldsymbol{S}\right\|_{\mathrm{F}}^2 < T_2 \quad \wedge \quad |\alpha_{\mathrm{R}} - \tilde{\alpha}| < T_3, \tag{35}$$

where $\boldsymbol{I}$ is the identity matrix, and T_2 and T_3 are preset thresholds, respectively. All pairs in $\mathbb{P}^*_{\tilde{\alpha}}$ are classified into inliers or outliers by checking the condition

$$\left\|(i,j)^\top - \boldsymbol{M} \cdot (k,l)^\top - \boldsymbol{t}\right\| < T_4, \tag{36}$$

for another threshold T_4. This procedure is repeated N_{iter} times, each time initialized with a triple of block pairs randomly drawn from set $\mathbb{P}^*_{\tilde{\alpha}}$. The algorithm outputs the set of pairings with the largest number of inliers as duplicated region.

Algorithm 1 Error Reduction Procedure

$\mathbb{P} = \{$Every matched block pair found by **LSH**$\}$
$\mathbb{R} = \{\ \}$ //result
for $\tilde{\alpha} = 0$ to 360 **do**
  $\mathbb{P}_{\tilde{\alpha}} = \{$Pairs from $\mathbb{P}$ that satisfy $|\alpha_{\mathrm{Z}} - \tilde{\alpha}| < T_1\}$
  $\mathbb{P}^*_{\tilde{\alpha}} = \{$Pairs from $\mathbb{P}_{\tilde{\alpha}}$ for which at least one spatially adjacent block pair is also included in $\mathbb{P}_{\tilde{\alpha}}\}$
  //Start RANSAC
  for iter $= 1$ **to** N_{iter} **do**
    Randomly choose three adjacent colinear pairs $(\boldsymbol{z}^{(i,j)}, \boldsymbol{z}^{(k,l)})$ from $\mathbb{P}^*_{\tilde{\alpha}}$
    //Estimate affine transformation
    Find $\boldsymbol{M}$ and $\boldsymbol{t}$ from $(i,j)^\top = \boldsymbol{M} \cdot (k,l)^\top + \boldsymbol{t}$
    //Singular Vector Decomposition
    $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$
    //Polar Decomposition
    $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top$, $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{V}^\top$
    $\alpha_{\mathrm{R}} = \arccos R_{11}$
    if $\|\boldsymbol{I} - S_{11}^{-1}\boldsymbol{S}\|_{\mathrm{F}}^2 \geq T_2$ or $|\alpha_{\mathrm{R}} - \tilde{\alpha}| \geq T_3$ **then**
      continue
    end if
    $$\mathbb{R}^*_{\tilde{\alpha}} = \left\{ \begin{matrix} \text{Pairs } \left(\boldsymbol{z}^{(i,j)}, \boldsymbol{z}^{(k,l)}\right) \text{ from } \mathbb{P}^*_{\tilde{\alpha}} \text{ satisfying} \\ \|(i,j)^\top - \boldsymbol{M} \cdot (k,l)^\top - \boldsymbol{t}\| < T_4 \end{matrix} \right\}$$
    if $|\mathbb{R}^*_{\tilde{\alpha}}| > |\mathbb{R}|$ **then**
      update $\mathbb{R}$ with $\mathbb{R}^*_{\tilde{\alpha}}$
    end if
  end for
end for

D. Complexity Analysis

This section analyzes the time complexity of the proposed method. Computing N_{moments} Zernike polynomials using (3) roughly takes $\mathcal{O}(N_{\mathrm{moments}})$. Determining actual Zernike moments is a point-wise multiplication of the corresponding polynomial and a block's intensity values. It is of order $\mathcal{O}(L^2)$ for a single moment. About $\mathcal{O}(N_{\mathrm{blocks}} \times N_{\mathrm{moments}} \times L^2)$ time is thus required to quantify all moments. The next components to

consider are locality sensitive hashing and searching the hashed data. Hashing approximately takes $\mathcal{O}(N_{\text{blocks}} \times N_{\text{moments}} \times P \times Q)$, where P and Q are the number of hash functions and buckets, respectively. Finding the nearest neighbor for all blocks requires

$$\mathcal{O}(N_{\text{blocks}} \times N_{\text{moments}} \times P \times Q) + \mathcal{O}(N_{\text{blocks}} \times N_{\text{moments}} \times \#\text{collisions}),$$

where #collisions is the number of near-duplicate vectors found in the hash tables using linear search. As L, P, and Q are relatively small, $\mathcal{O}(N_{\text{blocks}} \times N_{\text{moments}} \times L^2)$ takes similar time as $\mathcal{O}(N_{\text{blocks}} \times N_{\text{moments}} \times P \times Q)$. However, #collisions is rather large compared to the other parameters.[2] Hence, the time complexity up to the feature matching step is

$$\mathcal{O}(N_{\text{blocks}} \times N_{\text{moments}} \times \#\text{collisions}). \tag{37}$$

It depends, for a fixed image size and number of features, on the number of hash collisions, and thus on the image content. As for the time complexity of the error reduction procedure, we note that most false matches are already eliminated by (35) in practice. Despite two nested loops, the complexity of Algorithm 1 is consequently relatively low compared to feature extraction and matching, and we do not consider the time consumed by this last step any further.

V. Experimental Results

This section reports results from an extensive series of C(R)M localization and detection experiments. The setup includes our method as detailed in Section IV and an alternative detector, which combines Zernike moments with Christlein *et al.*'s k-d tree block matching and Same Affine Transformation Selection (SATS) framework [13].[3] We use the same framework for a number of further benchmark methods, namely the circular block method by Wang *et al.* [26], as well as Bashar *et al.*'s detectors based on discrete wavelet transform (DWT) and kernel PCA (KPCA) features [16]. The set of block-based detectors is completed by Bravo-Solorio and Nandi's [24] intensity-based algorithm. We also detail results of Pan and Lyu's SIFT-based approach [29] and report the performance of Amerini *et al.*'s SIFT-based detector [30] whenever appropriate. Note that all benchmark methods are particularly recommended for the localization and detection of rotated duplicated regions in Christlein *et al.*'s recent CM detection evaluation [43].

A. Experimental Setup

We use a randomly chosen subset of the BOSS image database [44] for our experiments. The image contents range from overly smooth to highly textured, making this data set a useful reference for benchmarking C(R)M detectors. Because creating convincing image manipulations is a time-consuming *manual* task, we follow prior research work and opt for an automated approach to generate a large number of processed images in a controlled environment. More specifically, we generate 1000 'manipulated' images by duplicating a randomly chosen part of the original image and inserting a possibly rotated version at a random[4] position of the same image without any local postprocessing. Rotation is implemented using bicubic interpolation. We take signal characteristics of the duplicated region into account and always consider each 500 'textured' and 'smooth' manipulations, whereby duplicate regions with $\geq$50 SIFT key points are assigned to the 'textured' set.[5] Duplication of textured regions is presumably easier to detect, because more distinct feature characteristics make block or key point matching more reliable. We believe that this distinction contributes to a better understanding of the strengths and weaknesses of different detection approaches and that it fosters a more comprehensive assessment of localization and detection performances.

If not stated otherwise, we use the the following parameters:

- The block size is set to $L = 24$. Zernike moments up to the of order $\bar{n} = 5$ are analyzed, which we found to give a good trade-off between the features' information content and sensitivity to noise [31] (see also Section IV-A). This results in feature vectors of length $N_{\text{moments}} = 12$.
- Locality sensitive hashing (cf. Section IV-B) is implemented based on the E2LSH package [45]. We found empirically that a distance threshold $D_1 = 300$ gives good results [31]. With preset D_1, a quantization bin size $w = 5000$ implies a collision probability of $\gamma_h(D_1) \geq 0.95$ by (26). For a success probability $\sigma = 0.9$, we set the length of hash vectors to $P = 30$ (and thus, by (28), $Q = 9$). Smaller P values are prone to produce many false alarms. The minimum spatial distance threshold between two matched blocks is set to $D_2 = 50$.
- The parameters of our phase-based error reduction procedure (cf. Section IV-C) have been determined empirically as follows. We set the maximum rotation angle differences $T_1 = T_3 = 2$. Threshold $T_2 = 0.03$ is chosen rather low, since our main objective are rotated blocks. The RANSAC parameters $T_4 = 2$ and $N_{\text{iter}} = 100$ resemble literature settings [29].

Benchmark methods ran with standard parameters as provided by the authors and/or suggested in the literature.[6]

Under these conditions, we evaluate localization and detection performance along four main directions.

- The *baseline experiment* (Section V-C) keeps both the image size and the size of duplicated regions constant, while systematically varying the rotation angle of the duplicated region, resulting in altogether 10000 runs per detector.
- *Variable-size experiments* (Section V-D) investigate the influence of the image size and the size of duplicated regions, respectively. Here, we rotate the duplicated region by a

[2]The expected number of collisions for a given query is $E[\#\text{collisions}] = Q \sum_{\boldsymbol{z} \in \mathbb{Z}} \gamma_h^P(\|\boldsymbol{z}^{(\text{query})} - \boldsymbol{z}\|)$.

[3]Christlein *et al.* [13] devised this combination as improvement of a detector presented by us in an earlier conference version of this manuscript [31].

[4]Overlap between original and duplicated region is ruled out by design.

[5]This definition is inspired by Pan and Lyu's experiments [29], who considered manipulations with $\geq$50 SIFT key points only.

[6]We set the SATS thresholds for the minimum Euclidean distance to 50 and for the minimum number of correspondences to 30, respectively. For Pan and Lyu's detector we set (using their symbols) the threshold for key point matching to $\epsilon = 0.5$, the maximum distance of RANSAC inliers to $\beta = 3$, and the number of RANSAC iterations to $N = 100$, respectively. Correlation maps are discretized with threshold $c = 0.3$.

Fig. 3. CRM manipulation and its detection result: (a) CRM manipulation involving a rotation of 10°; (b) duplicated region; and (c) detected region.

randomly chosen amount and vary either the size of the duplicated region or the size of the image. This series of experiments comprises in total 7000 runs per detector.

- *Robustness experiments* (Section V-E) address robustness against plausible postprocessing [46]. For this series of experiments, image size and size of the duplicated region are kept constant, whereas the rotation angle is chosen randomly. Overall 34,000 detector runs give insight how JPEG compression, distortion by additive white Gaussian noise, or blurring influence detection performance.
- *Copy—scale—move experiments* (Section V-F) examine localization and detection performance when the duplicated region underwent a rescaling operation instead of being rotated. While Zernike moments are not invariant to scaling by design, SIFT-based methods are also capable of handling arbitrary affine transformations other than rotation. Additional 6000 runs per detector thus further investigate strengths and weaknesses of the proposed method and the benchmark algorithms under this practically relevant scenario.

To the best of our knowledge, this large-scale setup accumulates to the most extensive evaluation of C(R)M detectors in the literature, which we further supplement by the analysis of a number of *real-life manipulations* in Section V-H.

B. Performance Metrics

We use pixel detection accuracy (PDA) and pixel false positive (PFP) rate for a quantitative evaluation of localization performance at *pixel-level*. These standard criteria [13], [19], [24], [29] reflect the ratio of correctly detected duplicated regions and the ratio of regions incorrectly marked as duplicate, respectively. PDA and PFP are defined as follows:

$$\mathrm{PDA} = \frac{\text{duplicate region} \cap \text{detected region}}{\text{duplicate region}} \tag{38}$$

$$\mathrm{PFP} = \frac{\text{detected region} - \text{duplicate region}}{\text{detected region}} \tag{39}$$

Fig. 3 visualizes the corresponding regions for an example manipulation. PDA and PFP are obtained by counting the respective pixels in the binary masks. In general, higher PDA as well as lower PFP values indicate superior performance.

While pixel-level metrics are useful to assess the general localization performance of C(R)M detectors when ground-truth data is available, *image-level* decisions are of particular interest to the automated detection of manipulated images. To this end, Section V-G reports ROC curves with image-level true positive rates (TPR) and false positive rates (FPR). Our detector uses the number of matched blocks as decision criterion, for which we vary the threshold from 0 to 20000 in steps of 20. ROC curves for the other detectors are drawn by thresholding the size of the detected region,[7] iterated over the sequence $\{0, 3^2, 6^2, \ldots, 513^2\}$. Since the design of Amerini *et al.*'s detector [30] (which adopts agglomerative hierarchical clustering on spatial locations) does not allow to compute ROC curves over the full FPR range in a straight-forward manner, we test the method as proposed in the original manuscript. We use Ward's linkage method and vary the clustering threshold T_h from 0.8 to 3.0 in steps of 0.2. FPR values are generally obtained by running the respective detector on all original images in the database.

C. Baseline Results

To demonstrate the general effectiveness of our detector, we generated each ten CRM manipulations from 1000 original images of size 512×512. Duplicated regions of size 96×96 were rotated in the range of 0° to 90°, applied in steps of 10°. Fig. 4 depicts selected box plots of PDA and PFP values for 500 textured (a) and 500 smooth (b) duplicated regions, as obtained with the proposed method and the six benchmark methods. Detailed localization results are also summarized in Tables I and II, respectively.

Taking all tested rotation angles into account, algorithms based on Zernike moments appear to be the most accurate approaches. At the same time, a direct comparison of our detector with the SATS algorithm [13] suggests that the proposed matching and error reduction procedure outperforms prior art especially for smooth duplicated regions. It is also worth noting that the proposed method generally yields lower PFP values than SATS. The impact of our feature space error reduction procedure (cf. Section IV-C) can be seen from Fig. 5, where we compare the baseline results of the proposed method with and without error reduction. Observe the drastically increased PFP values in the absence of pruning.

As for the other benchmark detectors in Fig. 4, we note that Wang *et al.*'s detector [26] gives a slightly better PDA, yet relatively higher PFP values impair the overall performance. Bravo-Solorio and Nandi's [24] intensity-based detector generally yields very low PFP values but fails to detect a large portion of duplicated regions for rotation angles other than 0° and 90°. Bashar *et al.*'s approaches [16] were not designed to be explicitly invariant to rotation and are thus not able to localize rotated regions at angles $\geq 10°$. Pan and Lyu's SIFT-based method [29] operates very well on the textured set, yet against the backdrop of relatively lower PDA values for smooth duplicated regions, where it is hard to extract a sufficient number of key points.

D. Variable-Size Results

Two important factors that impact the performance of C(R)M detectors are the size of the duplicated region and the size of the image under investigation [21]. While small duplicated regions are typically hard to distinguish from incorrectly matched blocks or key points, the likelihood of observing such false positives is expected to increase with the overall image size.

[7]The benchmark implementations worked best based on the region size and/or did not support counting the number of matched blocks.

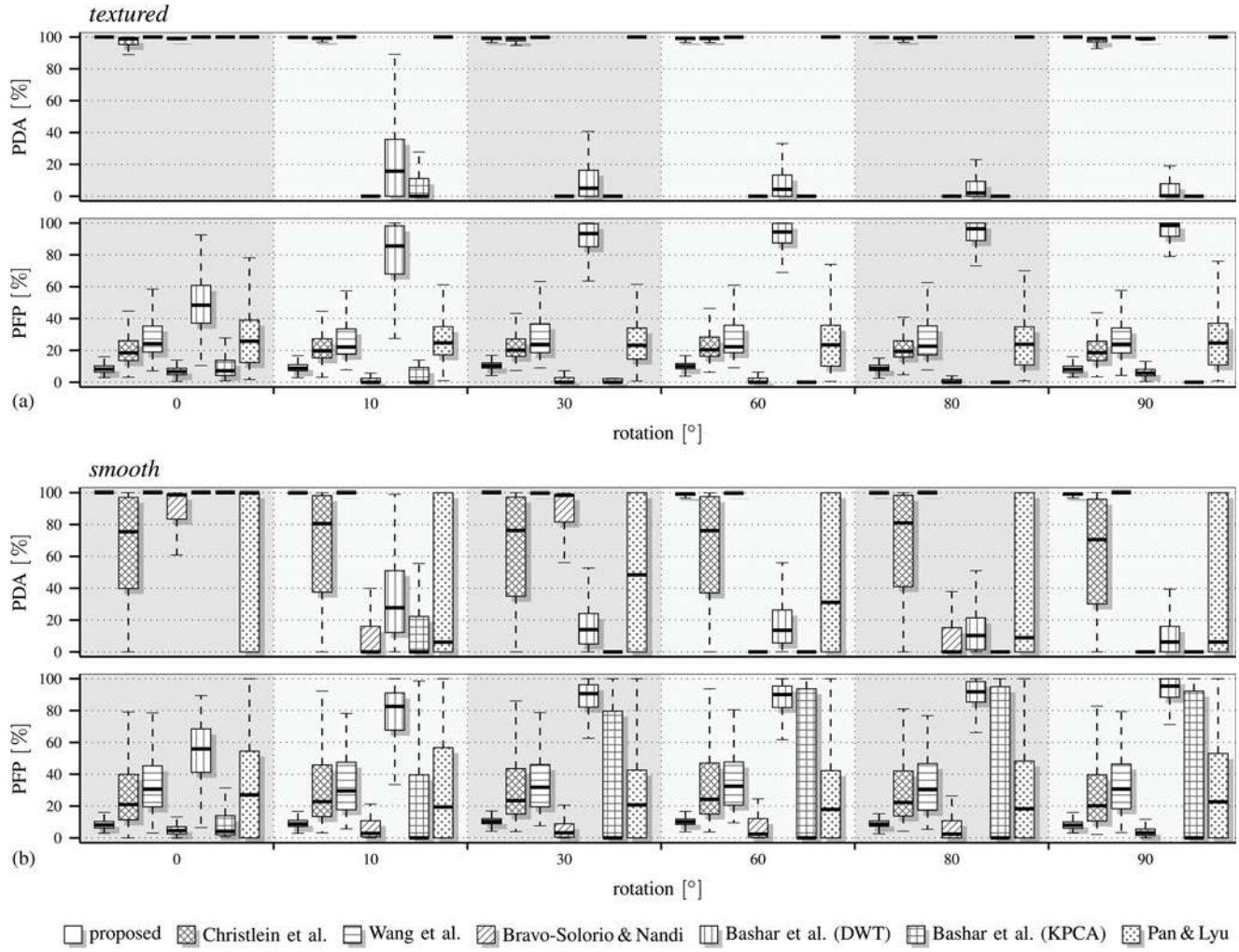

Fig. 4. Pixel-level CRM localization baseline results; textured (top rows) and smooth (bottom rows) subset PDA and PFP box plots; image size 512 × 512; duplicated region size 96 × 96; each 500 CRM manipulations per rotation angle.

TABLE I
TEXTURED SUBSET BASELINE PIXEL-LEVEL MEDIAN PDA AND PFP VALUES; BREAKDOWN BY DETECTOR AND ROTATION ANGLE

	median PDA [%] / PFP [%]						
	Zernike moments		intensity		frequency	dimensionality reduction	SIFT
rotation	proposed	Christlein et al.	Wang et al.	Bravo-Solorio & Nandi	Bashar et al. (DWT)	Bashar et al. (KPCA)	Pan & Lyu
0°	100.0 / 9.5	98.9 / 18.5	100.0 / 24.2	99.1 / 6.7	100.0 / 48.4	100.0 / 7.2	100.0 / 25.8
10°	99.7 / 9.3	99.4 / 19.7	100.0 / 22.1	0.0 / 0.0	15.7 / 85.6	0.0 / 0.0	100.0 / 24.7
20°	99.4 / 9.8	99.4 / 20.0	100.0 / 24.5	0.0 / 0.0	4.1 / 94.7	0.0 / 0.0	100.0 / 23.5
30°	98.6 / 9.9	99.3 / 20.2	99.9 / 23.8	0.0 / 0.0	5.0 / 93.4	0.0 / 0.0	100.0 / 23.2
40°	97.9 / 9.9	99.2 / 20.2	99.9 / 23.5	0.0 / 0.0	2.2 / 96.8	0.0 / 0.0	100.0 / 23.3
50°	97.7 / 10.0	99.2 / 20.1	99.9 / 23.9	0.0 / 0.0	2.8 / 96.5	0.0 / 0.0	100.0 / 22.9
60°	98.5 / 9.9	99.4 / 20.4	99.9 / 22.2	0.0 / 0.0	4.3 / 94.4	0.0 / 0.0	100.0 / 23.5
70°	99.4 / 10.2	99.5 / 21.0	100.0 / 21.9	0.0 / 0.0	3.1 / 95.5	0.0 / 0.0	100.0 / 23.4
80°	99.7 / 9.1	99.5 / 19.4	100.0 / 22.6	0.0 / 0.0	2.0 / 96.4	0.0 / 0.0	100.0 / 23.9
90°	100.0 / 9.7	99.1 / 18.6	100.0 / 23.9	99.0 / 5.8	0.2 / 98.5	0.0 / 0.0	100.0 / 24.6
total	99.4 / 9.7	99.3 / 19.8	100.0 / 23.3	0.0 / 2.4	5.2 / 93.5	0.0 / 0.0	100.0 / 23.9

We examine the influence of the duplicated region size by increasing this dimension from 32 × 32 to 128 × 128 pixels in steps of 32 square pixels while keeping the image size (512 × 512) constant. For each size, 500 textured and 500 smooth CRM manipulations were created by rotating the duplicated region by an angle randomly chosen from the set $\{0°, 10°, \ldots, 90°\}$. Following our definition of smoothness, duplicated regions of size 32 × 32, 64 × 64, 96 × 96, and 128 × 128 with more than 6, 25, 50, and 90 SIFT key points, respectively, are labeled 'textured', thus keeping the relative number of necessary key

TABLE II
SMOOTH SUBSET BASELINE PIXEL-LEVEL MEDIAN PDA AND PFP VALUES; BREAKDOWN BY DETECTOR AND ROTATION ANGLE

	median PDA [%] / PFP [%]						
	Zernike moments		intensity		frequency	dimensionality reduction	SIFT
rotation	proposed	Christlein et al.	Wang et al.	Bravo-Solorio & Nandi	Bashar et al. (DWT)	Bashar et al. (KPCA)	Pan & Lyu
0°	100.0 / 8.2	75.5 / 20.9	100.0 / 30.6	98.6 / 4.6	100.0 / 55.9	100.0 / 4.2	99.8 / 27.0
10°	99.8 / 8.8	80.6 / 22.8	99.9 / 29.5	0.0 / 2.9	27.7 / 82.6	0.0 / 0.0	6.1 / 19.4
20°	99.7 / 10.0	76.1 / 22.9	99.8 / 29.6	0.0 / 3.3	9.7 / 92.5	0.0 / 0.0	5.9 / 17.9
30°	99.3 / 10.4	76.2 / 23.4	99.7 / 31.8	0.0 / 3.2	14.0 / 90.6	0.0 / 0.0	31.1 / 20.7
40°	99.0 / 10.3	71.7 / 23.4	99.6 / 32.1	0.0 / 3.1	8.9 / 93.4	0.0 / 0.0	12.3 / 18.8
50°	99.1 / 10.3	76.5 / 22.8	99.7 / 32.6	0.0 / 5.1	8.1 / 93.4	0.0 / 0.0	24.2 / 20.6
60°	99.2 / 10.1	70.5 / 24.2	99.7 / 32.5	0.0 / 2.4	13.6 / 90.1	0.0 / 0.0	6.2 / 17.9
70°	99.6 / 9.8	77.2 / 23.9	99.8 / 32.0	0.0 / 1.7	9.1 / 93.2	0.0 / 0.0	3.6 / 17.3
80°	99.8 / 8.7	81.0 / 22.3	99.9 / 30.3	0.0 / 2.4	10.3 / 91.8	0.0 / 0.0	8.9 / 18.3
90°	100.0 / 8.0	76.3 / 20.2	100.0 / 30.7	98.4 / 3.2	6.2 / 95.4	0.0 / 0.0	48.4 / 22.7
total	99.6 / 9.5	76.1 / 22.9	99.8 / 31.3	0.0 / 3.5	13.1 / 90.4	0.0 / 0.0	14.6 / 20.3

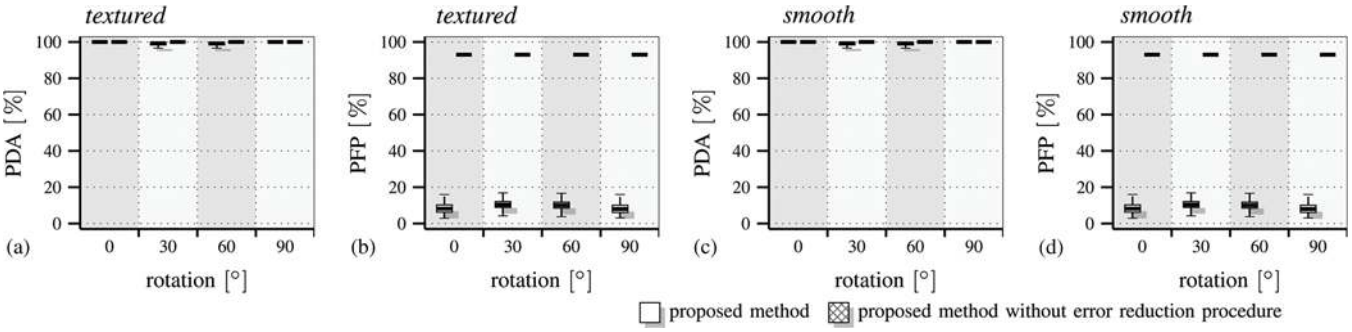

Fig. 5. Pixel-level CRM localization baseline results with/without error reduction procedure for the proposed method; textured (left) and smooth (right) subset PDA and PFP box plots; image size 512 × 512; duplicated region size 96 × 96; each 500 CRM manipulations per rotation angle.

points per region size constant. The left panels of Fig. 6 report box plots of the resulting pixel-level PDA values[8] for textured (a) and smooth (c) regions. While the results from both subsets generally resemble the detectors' baseline performances, it is worth noting that all but Christlein *et al.*'s [13] and Wang *et al.*'s [26] methods fail to reliably localize duplicated regions of size 32 × 32. Yet these methods suffer from relatively high median PFP values (54.3/50.7% and 53.9%/79.7% for textured/smooth 32 × 32 regions, respectively). Hence, we conclude in accordance to earlier reports in the literature [29] that duplicated regions of this size are typically too small for state-of-the-art CRM detectors.

In addition to the duplicated region size, we further investigated the influence of image size by duplicating a randomly-chosen, fixed-size region (96 × 96) in images of dimension 256 × 256, 512 × 512, and 1024 × 1024, respectively. As before, the duplicated regions (each 500 textured and 500 smooth) were rotated by a randomly chosen angle. Note that the 256 × 256 images were generated by down-sampling the corresponding BOSS images by a factor of two, whereas the 1024 × 1024 images were resized and cropped from the BOSS RAW database. Fig. 6(b) and (d) depict the resulting textured and smooth subset PDA box plots for each image size. The graphs indicate that our method, along with Wang *et al.*'s detector, works with high accuracy relatively independent the image size and the smoothness of duplicated regions. Observe that this is generally not the case for all other benchmark methods.

In summary, the variable-size experiments confirm the baseline results. Fig. 6 indicates that CRM detectors based on Zernike moments or circular features are often superior. Our detector is among the best performers for duplicated regions larger than 32 × 32, whereas the SIFT-based method localizes in particular textured duplicated regions with very high accuracy.

E. Robustness Results

Assessment of forensic algorithms' detection performance in practice also calls for considering their robustness against plausible postprocessing [46]. This concerns in particular lossy JPEG compression, which is often the last step prior to releasing a manipulation to the public. JPEG compression is likely to smooth out subtle differences between individual blocks. At the same time it introduces its own blocking artifacts. Here, we compress the image after CRM manipulation with JPEG quality factors 100, 80, 60, and 40, respectively. We further investigate robustness against additive white Gaussian noise (AWGN) and linear blurring with a circular averaging filter. We understand these image processing primitives as placeholders for more

[8]Due to space constraints, we refrain from reporting detailed PFP values in the remainder of the manuscript.

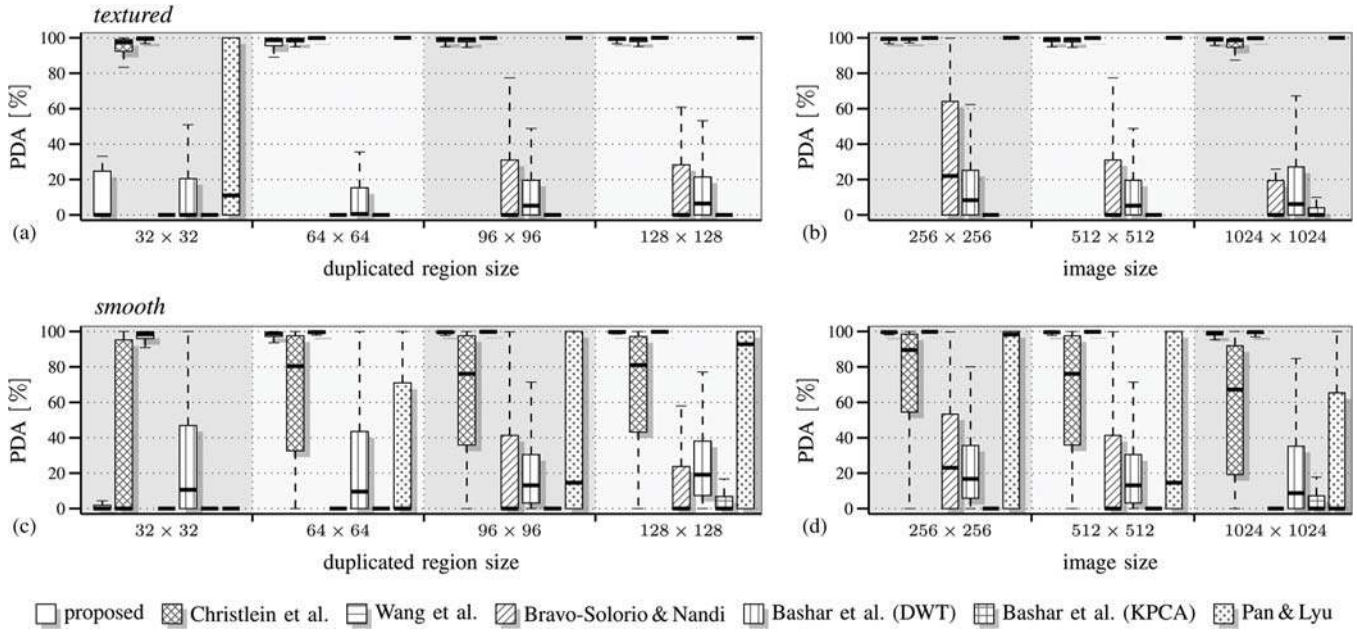

Fig. 6. Pixel-level CRM localization variable-size results; textured (top row) and smooth (bottom row) subset PDA box plots as obtained for (a, c) fixed-size 512 × 512 images with different duplicated region sizes, and (b, d) different image sizes with fixed-size 96 × 96 duplicated region size; each 500 CRM manipulations with rotation angles randomly sampled from the set $\{0°, 10°, \ldots, 90°\}$.

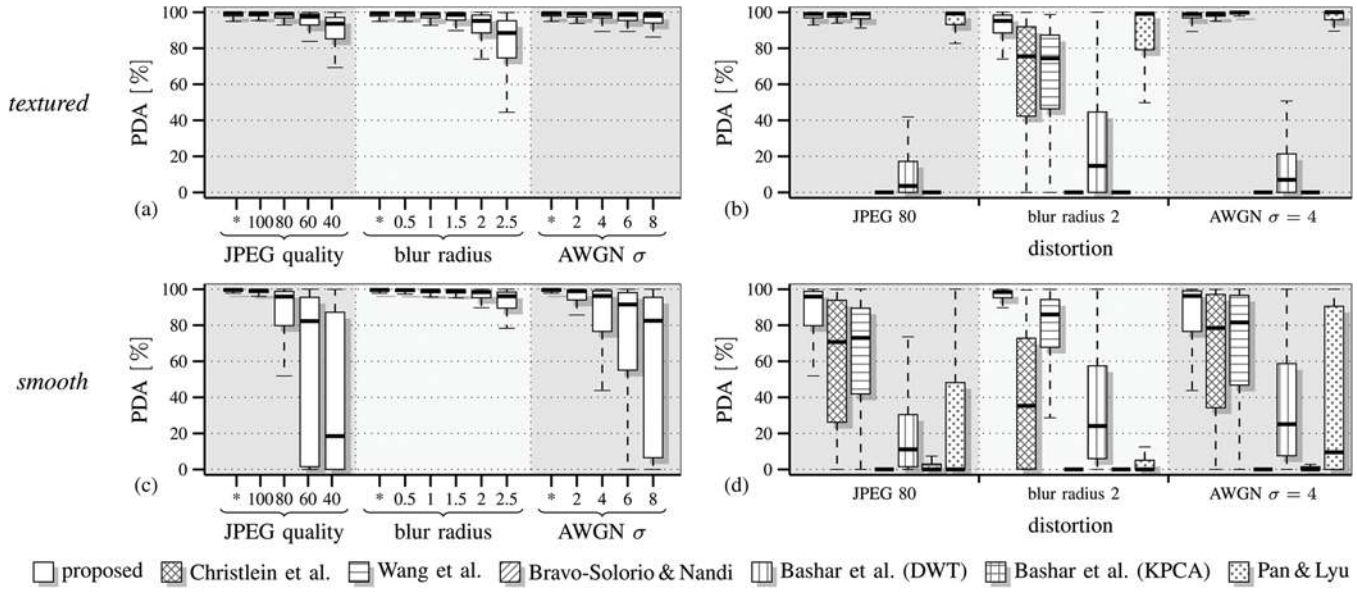

Fig. 7. Pixel-level CRM localization robustness results; textured (top row) and smooth (bottom row) subset PDA box plots as obtained (a, c) with our detector for different types and strengths of distortion, and (b, d) with all benchmark detectors for selected distortion settings; each 500 CRM manipulations with rotation angles randomly sampled from the set $\{0°, 10°, \ldots, 90°\}$; image size 512 × 512; duplicated region size 96 × 96; "∗" refers to no distortion.

complex manipulation procedures—which may also be part of intended targeted attacks [46] to hide characteristic CRM traces [47]—and apply them to the duplicated regions only. The strength of distortion is parametrized via the noise's standard deviation and the filter's radius. We considered zero-mean AWGN with standard deviation 2, 4, 6, and 8, as well as average filters of radius 0.5, 1, 1.5, 2, and 2.5, respectively. Note that all images in this series of experiments (each 500 from the textured set and 500 from the smooth set) are of size 512 × 512, with 96 × 96 duplicated regions rotated by randomly sampling from the set $\{0°, 10°, \ldots, 90°\}$. Under these settings, Fig. 7 summarizes the robustness of our detector and the benchmark approaches. More specifically, panels (a) and (c) report, for the textured set and the smooth set, respectively, PDA box plots of the proposed method for (from left to right) all tested JPEG quality factors, blur radii, and noise strengths. Overall, the results emphasize the high robustness of Zernike moments. The right panels in Fig. 7 compare all seven detectors for one representative JPEG, blur and noise parameter setting. The graphs indicate that the proposed method is superior to alternative approaches throughout all benchmark experiments for both textured (b) and smooth (d) duplicated regions. Observe that

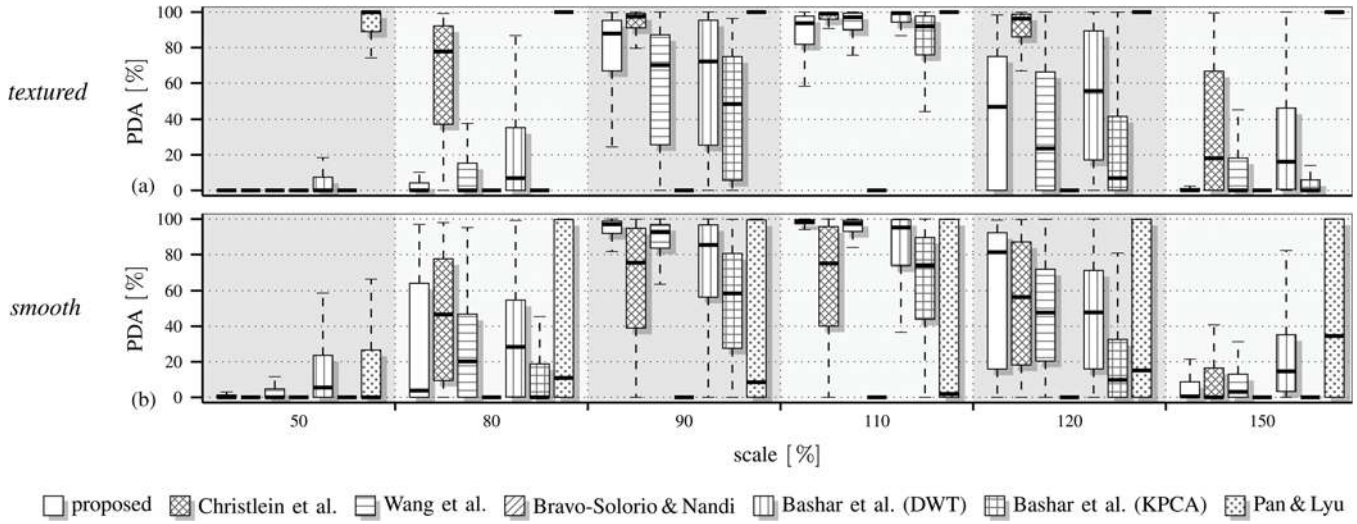

Fig. 8. Pixel-level copy—scale—move localization results; (a) textured and (b) smooth subset PDA box plots; image size 512 × 512; duplicated region size 96 × 96; each 500 CRM manipulations per scaling factor.

the performance of Wang *et al.*'s detector is affected relatively stronger, presumably because their approach directly depends on the blocks' intensity values.

F. Copy—Scale—Move Results

In a special case of robustness experiments, we also investigated localization performance when duplicated regions underwent geometric transformations other than rotation. We frame this as a question of robustness because Zernike moments are by design only invariant to rotation. Yet realistic copy—move manipulations may also incorporate other transformation types, so that we deem a reasonable localization accuracy even under this more general setting a valuable property of CRM detectors. At the same time, we see this experiment as a critical test of our detector with respect to SIFT-based methods, which can by definition also handle arbitrary affine transformations.

We generated six CM manipulations from each of the 500 images of size 512 × 512 in the textured set and in the smooth set. The original region size was 96 × 96, and the corresponding duplicated regions were rescaled by $\pm\{10, 20, 50\}\%$. Fig. 8 depicts the resulting PDA box plots as obtained with our detector and the six benchmark methods. The graphs indicate that our method remains accurate for slightly rescaled $(\pm 10\%)$ textured regions and even outperforms all other detectors when smooth regions are rescaled to [90, 120]%. Not surprisingly, performance degrades towards stronger scaling factors, so that the proposed method is not able to localize duplicated regions in the more extreme cases. For textured regions, the SIFT-based detector yields almost perfect PDA values, yet performance again strongly depends on the image content.

G. Image-Level Manipulation Detection Results

Figs. 9 and 10 proceed with an overview of the image-level manipulation detection performances. The individual subfigures resemble the previous localization experiments in Sections V-C, V-D, V-E and V-F. More specifically, Fig. 9 reports textured and smooth subset ROC curves of our detector and the six benchmark methods in the baseline experiment, the variable size experiments, and the robustness experiments. Corresponding textured and smooth region results for Amerini *et al.*'s detector [30] in Table III complement the curves. All results were obtained according to the procedure outlined in Section V-B. Each data point is based on 1000 original images and 500 textured/smooth CRM manipulations with rotation angles randomly sampled from $\{0°, 10°, \ldots, 90°\}$. Table IV summarizes the results for all methods at a fixed 4.1% FPR.[9] For the sake of a fair comparison, we also report the detection performance of Pan & Lyu's method with the threshold mentioned in their paper (0.1% of total image size, corresponding to a FPR of 11.1%). Fig. 10 presents the textured and smooth subset copy—scale—move ROC curves for selected scaling factors. Table V completes the benchmark with results of Amerini *et al.*'s detector under this scenario.

Overall, the image-level detection results are largely in line with the pixel-level localization performances. For smooth regions, the proposed method outperforms all benchmark detectors in every tested circumstance except for strong scaling, cf. Fig. 10(e) and (h). SIFT-based methods are particularly well suited for detecting textured duplicated regions, for which Christlein *et al.*'s Zernike/SATS combination shows a similar performance (except for strong scaling, cf. Fig. 10(a) and (d)). It is worth noting that all six block-based detectors suffer from rather low detection rates at very low FPR values, whereas our method tends to give better results in this particular segment of the ROC curve (cf. Table IV). Wang *et al.*'s detector generally gives higher FPR's than our method in Fig. 9. Bravo-Solorio and Nandi's detector typically cannot compete with the other methods. Very strong scaling impairs the performance of all detectors, most apparently when smooth duplicated regions are concerned (cf. Fig. 10(e) and (h), and Table V). In general, we believe that *all* ROC curves leave room for improvement.

[9]For Amerini *et al.*'s detector, this corresponds to a clustering threshold $T_h = 2.4$, which we found to give the best results under most settings (cf. Table III).

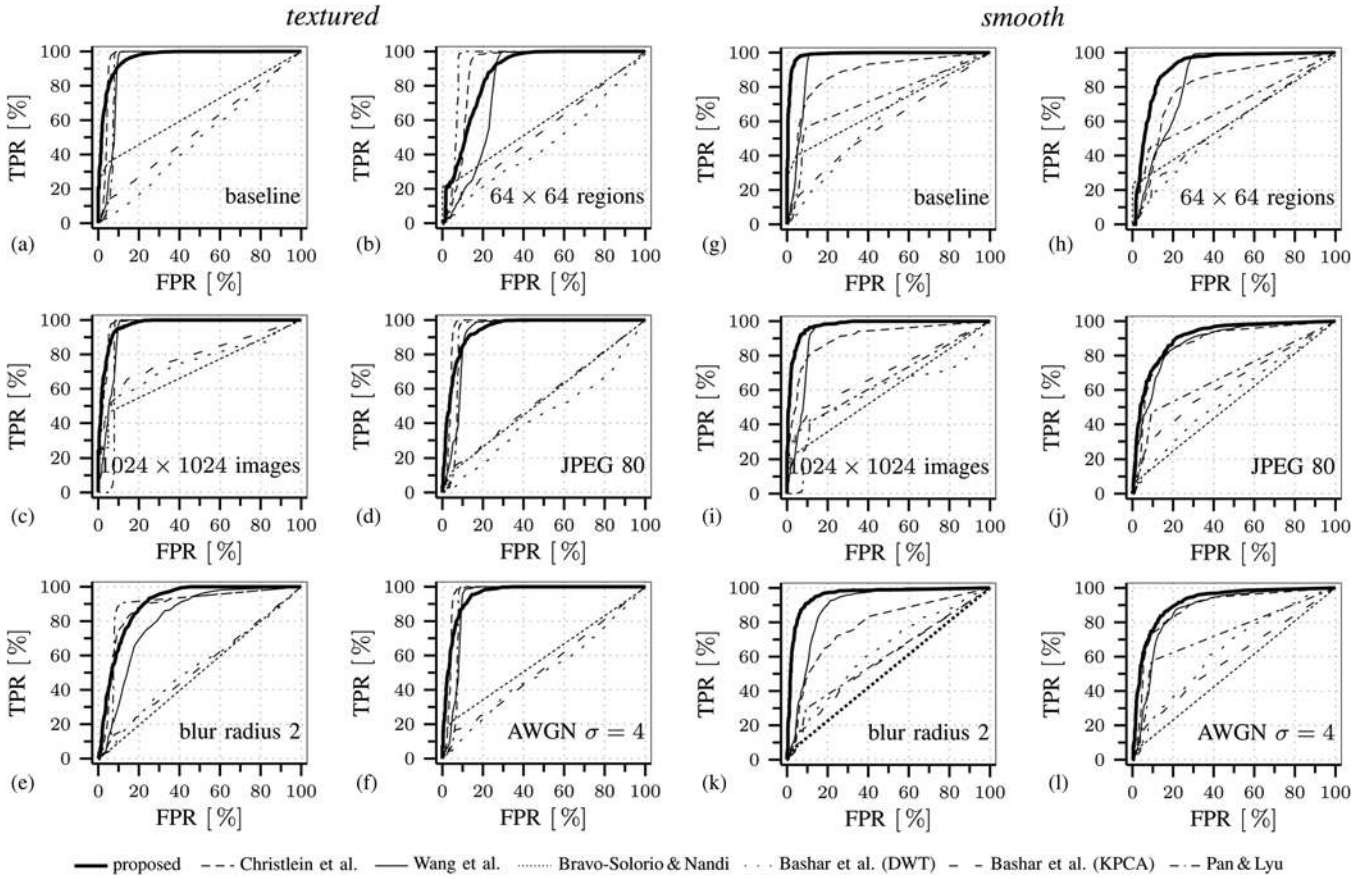

Fig. 9. Textured and smooth subset image-level CRM detection; ROC curves from 1000 originals and each 500 manipulations with rotation angles randomly sampled from the set $\{0°, 10°, \ldots, 90°\}$; (a, g) duplicated region size 96×96, image size 512×512, (b, h) duplicated region size 64×64, (c, i) image size 1024×1024, (d, j) JPEG compression with QF 80, (e, k) blur with radius 2.0, and (f, l) AWGN with $\sigma = 4$.

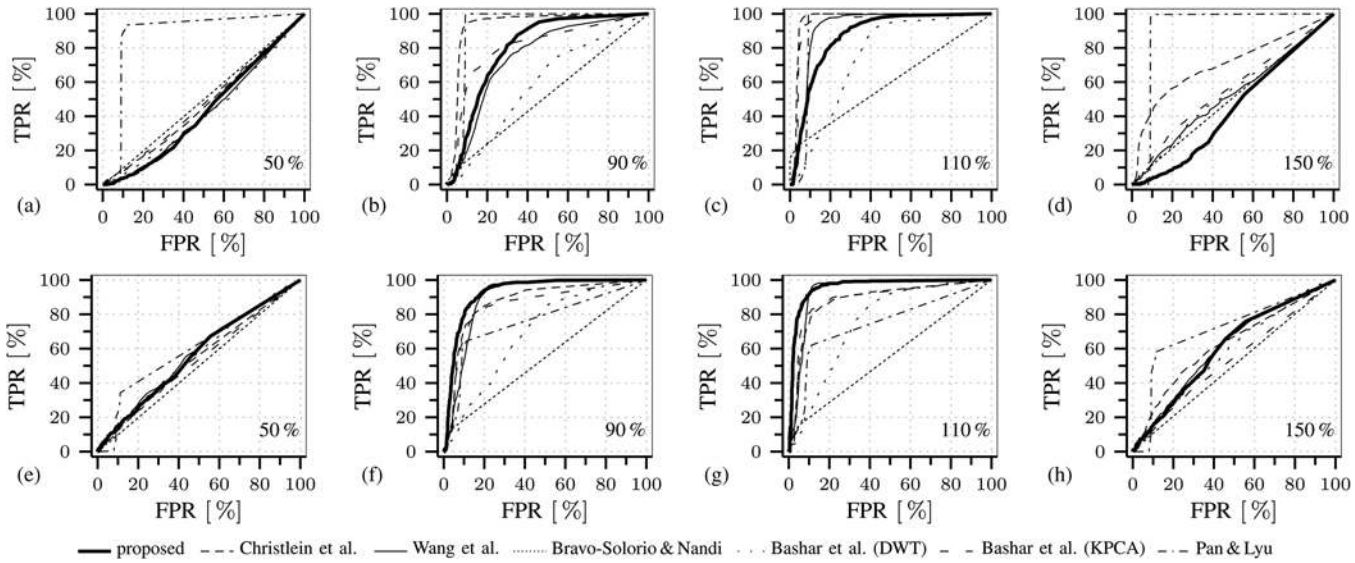

Fig. 10. Image-level copy—scale—move detection results for textured (top row) and smooth (bottom row) duplicated regions; ROC curves from 1000 original images and each 500 CRM manipulations; 96×96 duplicated regions rescaled to (a, e) 50%, (b, f) 90%, (c, g) 110%, and (d, h) 150%.

H. Tests With Real-Life Manipulations

Our last series of experiments analyzes five realistic manipulations of the Erlangen 'Image Manipulation Dataset', which Christlein *et al.* [43] compiled for the purpose of benchmarking CM detectors. The images cover different levels of sophistication, with duplicated regions of varying number and size.[10]

Fig. 11 depicts two images ('acropolis' and 'giraffe') along with ground truth masks in Fig. 11(b). The output of our detector

[10]http://www5.cs.fau.de/research/data/image-manipulation/

TABLE III
TEXTURED AND SMOOTH SUBSET IMAGE-LEVEL CRM DETECTION WITH AMERINI'S METHOD AT DIFFERENT CLUSTERING THRESHOLDS T_h; BASELINE RESULTS, ALONG WITH SELECTED VARIABLE SIZE AND ROBUSTNESS RESULTS; 1000 ORIGINALS AND EACH 500 MANIPULATIONS

		True Positive Rate (textured / smooth) [%]					
T_h	FPR [%]	baseline	region size 64×64	image size 1024×1024	JPEG 80	blur radius 2.0	AWGN $\sigma = 4$
0.8	0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
1.0	0.0	2.3 / 0.3	1.4 / 0.0	2.6 / 0.0	0.6 / 0.0	0.2 / 0.0	0.2 / 0.0
1.2	0.5	29.9 / 4.0	14.8 / 2.4	32.8 / 1.0	19.6 / 2.8	7.8 / 1.4	21.8 / 1.0
1.4	1.7	83.0 / 18.1	57.8 / 14.2	82.4 / 5.4	62.6 / 12.0	37.6 / 7.8	77.6 / 10.0
1.6	3.4	99.5 / 38.1	91.6 / 36.0	99.2 / 13.6	90.0 / 24.0	66.0 / 21.8	98.4 / 21.8
1.8	4.4	100.0 / 48.2	99.2 / 47.4	99.8 / 18.4	96.6 / 33.4	81.8 / 33.2	99.8 / 30.4
2.0	4.6	100.0 / 49.8	99.6 / 49.8	100.0 / 19.2	97.6 / 34.4	84.6 / 34.6	100.0 / 33.8
2.2	4.5	100.0 / 51.2	100.0 / 50.6	100.0 / 19.6	98.2 / 34.8	86.4 / 35.6	100.0 / 35.6
2.4	4.1	100.0 / 51.2	100.0 / 51.0	100.0 / 21.4	98.4 / 35.0	86.4 / 36.4	100.0 / 36.2
2.6	4.0	100.0 / 50.8	100.0 / 50.2	100.0 / 19.0	97.8 / 34.4	86.8 / 36.0	100.0 / 35.2
2.8	3.0	100.0 / 48.6	99.4 / 48.0	99.8 / 16.4	97.4 / 32.2	84.6 / 31.8	100.0 / 33.0
3.0	1.7	99.9 / 42.7	97.2 / 40.0	98.8 / 10.0	95.8 / 26.4	75.6 / 25.8	99.6 / 26.0

TABLE IV
IMAGE-LEVEL TPR AT FPR $= 4.1\%$; 1000 ORIGINALS, 500 TEXTURED AND 500 SMOOTH CRM MANIPULATIONS

		True Positive Rate [%]	
		textured	smooth
Zernike moments	proposed	75.7	95.5
	Christlein et al.	49.0	25.7
intensity	Wang et al.	12.7	26.0
	Bravo-Solorio & Nandi	33.8	38.5
frequency	Bashar et al. (DWT)	2.9	4.7
Dim. reduction	Bashar et al. (KPCA)	10.8	9.8
SIFT	Pan & Lyu (FPR: 4.1%)	7.0	13.2
	Pan & Lyu (FPR: 11.1%)	100.0	57.5
	Amerini et al.	100.0	51.2

TABLE V
TEXTURED AND SMOOTH SUBSET IMAGE-LEVEL COPY—SCALE—MOVE DETECTION WITH AMERINI'S METHOD AT DIFFERENT CLUSTERING THRESHOLDS T_h; 1000 ORIGINALS AND EACH 500 MANIPULATIONS

		True Positive Rate (textured / smooth) [%]			
T_h	FPR [%]	scale 50%	scale 90%	scale 110%	scale 150%
0.8	0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
1.0	0.0	0.0 / 0.0	1.2 / 0.0	0.6 / 0.6	1.8 / 0.0
1.2	0.5	6.2 / 1.6	22.2 / 2.8	25.4 / 3.6	28.4 / 5.2
1.4	1.7	34.0 / 5.8	77.2 / 16.4	85.2 / 16.4	80.4 / 17.2
1.6	3.4	70.0 / 14.0	98.4 / 33.2	99.6 / 33.0	97.8 / 33.6
1.8	4.4	86.2 / 22.0	100.0 / 46.6	100.0 / 43.0	99.4 / 44.6
2.0	4.6	89.6 / 22.4	100.0 / 48.2	100.0 / 45.8	99.6 / 47.2
2.2	4.5	91.2 / 23.4	100.0 / 50.0	100.0 / 47.8	99.6 / 48.6
2.4	4.1	91.2 / 23.6	100.0 / 50.4	100.0 / 47.6	99.6 / 49.0
2.6	4.0	91.0 / 23.0	100.0 / 49.8	100.0 / 47.2	99.6 / 48.0
2.8	3.0	89.6 / 21.2	100.0 / 46.6	100.0 / 45.2	99.6 / 45.8
3.0	1.7	83.0 / 14.0	99.8 / 39.2	100.0 / 38.4	99.4 / 36.6

is displayed in Fig. 11(c). It coincides to a great extent with the true duplicated regions. Table VI summarizes the detection results for all five (partly downscaled) tested images and compares our detector to the benchmark methods. The PDA/PFP values largely reflect the large-scale results of the previous sections. Except for the 'building' image (with a duplicated region smaller than 32×32), all detectors are generally capable

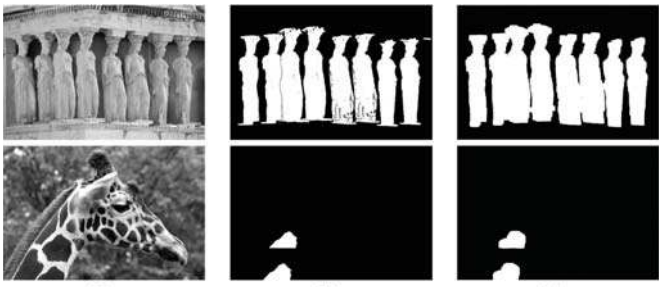

Fig. 11. "Acropolis" and "giraffe" CM manipulation [43]: (a) manipulated images, (b) duplicated regions, and (c) detected regions.

of localizing the duplicated regions. Overall, Zernike moments offer one of the best trade-offs between high PDA and relatively low PFP. Compared to Christlein *et al.*'s SATS algorithm [13], our feature-based error reduction procedure tends to give lower PFP's at similar PDA values. Similarly, the proposed method also yields lower PFP values than Wang *et al.* [26], Bashar *et al.* (DWT) [16], and Pan and Lyu [29]. Bashar *et al.*'s KPCA detector [16] works very well for most images, mainly because the duplicated regions were not rotated. Bravo-Solorio and Nandi's method [24] fails to detect large portions of the 'acropolis' manipulation.

I. Summary

All experimental results strongly suggest that localization and detection performance greatly depends on image characteristics, the type of processing applied to the duplicated region, and many other influencing factors. This emphasizes both the empirical nature and the complexity of the problem, and it calls for a careful interpretation of detector outputs in practical investigations. In practice, also a detector's run-time has to be considered. Table VII reports average run times of the tested detectors, obtained from the analysis of each 1000 images of size 512×512 with standard parameter settings. Among the tested block-based methods, our detector is fastest, outperforming others by a factor of 2 at least. Not surprisingly, the table also finds that all these detectors are considerably slower than SIFT-based methods. Yet the high efficiency of SIFT-based detectors comes at the cost

TABLE VI
PIXEL-LEVEL PDA AND PFP FOR FIVE REALISTIC CM MANIPULATIONS

		PDA [%] / PFP [%]						
		Zernike moments		intensity		frequency	dimensionality reduction	SIFT
image	size	proposed	Christlein et al.	Wang et al.	Bravo-Solorio & Nandi	Bashar et al. (DWT)	Bashar et al. (KPCA)	Pan & Lyu
giraffe	800 × 533	98.4 / 27.4	98.5 / 38.1	97.5 / 32.6	91.2 / 23.1	96.9 / 82.2	96.3 / 25.7	100.0 / 95.3
tree	1024 × 683	99.4 / 61.8	99.8 / 61.5	96.8 / 64.1	83.2 / 38.3	99.5 / 77.4	98.5 / 59.2	100.0 / 86.0
cattle	1280 × 854	87.1 / 36.6	99.1 / 44.3	96.4 / 59.4	86.6 / 24.0	93.3 / 30.2	93.2 / 26.9	74.2 / 96.5
acropolis	1024 × 685	98.6 / 12.8	77.9 / 22.9	99.5 / 23.3	7.2 / 3.0	99.5 / 30.9	95.0 / 15.4	100.0 / 36.0
building	1024 × 809	1.7 / 99.9	100.0 / 98.8	100.0 / 95.7	0.0 / 0.0	100.0 / 99.3	100.0 / 98.5	61.3 / 99.2

TABLE VII
AVERAGE RUN TIME OF EACH DETECTOR FOR 1000 512 × 512 IMAGES

	block-based						SIFT-based	
	proposed	Christlein et al.	Wang et al.	Bravo-Solorio & Nandi	Bashar et al. (DWT)	Bashar et al. (KPCA)	Pan & Lyu	Amerini et al.
run-time	342 s	687 s	1115 s	1093 s	1215 s	1534 s	10 s	37 s

of suboptimal detection performance when the duplicated regions have not enough texture information. Overall, we thus conclude that the quest for the one "best" detector rather depends on the investigator's priorities of how computational resources, low false positives and high detection rates shall be traded off against each other.

VI. CONCLUDING REMARKS

In this paper, we have focused on the localization and detection of copy—rotate—move (CRM) manipulations, for which we advocate Zernike moments as robust feature representation of small overlapping image blocks. Zernike moments are known for their analytical invariance to rotation and are thus particularly suitable for this application. Our detector employs a novel block matching paradigm based on locality sensitive hashing, and it exploits phase differences of Zernike moments in a feature-based error reduction approach. Extensive experimental results based on large set of images confirm the superior performance and robustness under a variety of settings.

As to the limitations, we note that detectors based on Zernike moments are inherently incapable of localizing duplicated regions that underwent strong affine transformations other than rotation. In general, it remains an open research question whether there exists one single "catch-all" approach, or whether combining specialized detectors with feature representations invariant to particular forms of processing is more favorable. The signal-dependent performance of different CRM detectors, the relatively low computational cost of running individual state-of-the-art detection procedures in parallel, and recent decision fusion attempts [48] can all be viewed in support of the latter option. Moreover, it is also conceivable to support duplicated region detectors with side-information about other image characteristics, for instance traces of CFA interpolation [49] or JPEG compression [50].

Apart from endeavors to advance the field by means of more robust feature representations, we see a major domain for future research in the process of making automated decisions about the presence of C(R)M manipulations. Given a typical binary output map (see for instance Fig. 11(c)), human interpretation is likely to unveil the presence of suspicious regions with high accuracy by comparison with the questioned image. On the contrary, rather simple computational measures, such as plain integration over the output map, ignore available semantical information. They are thus particularly prone to false alarms. We see this as being part of a more general problem common to all forensic methods aiming at local image manipulations. It is up to future research to investigate means how recent advances in *quantifying* the 'semantical impact' of a manipulation in a nonblind setting [51] can be transferred to the examination of (localized) forensic detection scores with side information about the image content.

ACKNOWLEDGMENT

The authors thank I. Amerini *et al.* [30], S. Bravo-Solorio and A. K. Nandi [24], V. Christlein and C. Riess [13], [43], and P. Indyk [45] for sharing their CRM detection and LSH code.

REFERENCES

[1] H. Farid, Photo Tampering Throughout History 2013 [Online]. Available: http://www.fourandsix.com/photo-tampering-history
[2] , H. T. Sencar and N. Memon, Eds., *Digital Image Forensics: There is More to a Picture than Meets the Eye*. New York, NY, USA: Springer, 2013.
[3] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 758–767, Feb. 2005.
[4] M. Kirchner, "Linear row and column predictors for the analysis of resized images," in *Proc. ACM SIGMM Multimedia & Security Workshop*, New York, NY, USA, 2010, pp. 13–18, ACM Press.
[5] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Information Hiding*, ser. Lecture Notes in Computer Science, J. Fridrich, Ed. Berlin, Heidelberg, Germany: Springer-Verlag, 2004, vol. 3200, pp. 128–147.
[6] Y.-L. Chen and C.-T. Hsu, "Detecting recompression of JPEG images via periodicity analysis of compression artifacts for tampering detection," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 2, pp. 396–406, Jun. 2011.
[7] M. C. Stamm and K. J. R. Liu, "Forensic detection of image manipulation using statistical intrinsic fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 492–506, Sep. 2010.

[8] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," in *Media Forensics and Security II*, ser. Proceedings of SPIE, N. Memon, J. Dittmann, A. Alattar, and E. Delp, Eds. Bellingham, WA, USA: SPIE, 2010, vol. 7541, p. 754110.
[9] G. Cao, Y. Zhao, R. Ni, and A. C. Kot, "Unsharp masking sharpening detection via overshoot artifacts analysis," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 603–606, Oct. 2011.
[10] R. Böhme, F. Freiling, T. Gloe, and M. Kirchner, "Multimedia forensics is not computer forensics," in *Computational Forensics*, ser. Lecture Notes in Computer Science, Z. J. Geradts, K. Y. Franke, and C. J. Veenman, Eds. Berlin, Heidelberg, Germany: Springer-Verlag, 2009, vol. 5718, pp. 90–103.
[11] J. Fridrich, D. Soukal, and J. Lukáš, "Detection of copy-move forgery in digital images," in *Proc. Digital Forensic Research Workshop*, Cleveland, OH, USA, 2003.
[12] M. R. Teague, "Image analysis via the general theory of moments," *J. Opt. Soc. Amer.*, vol. 70, no. 8, pp. 920–930, 1980.
[13] V. Christlein, C. Riess, and E. Angelopoulou, "On rotation invariance in copy-move forgery detection," in *Proc. IEEE Int. Workshop on Information Forensics and Security*, Seattle, WA, USA, 2010, pp. 1–6.
[14] Y. Huang, W. Lu, W. Sun, and D. Long, "Improved DCT-based detection of copy-move forgery in images," *Forensic Sci. Int.*, vol. 206, no. 1–3, pp. 178–184, 2011.
[15] A. C. Popescu and H. Farid, Exposing Digital Forgeries by Detecting Duplicated Image Regions, Computer Science, Dartmouth College, Tech. Rep., 2004.
[16] M. Bashar, K. Noda, N. Ohnishi, and K. Mori, "Exploring duplicated regions in natural images," *IEEE Trans. Image Process.*, to be published.
[17] J.-W. Wang, G.-J. Liu, Z. Zhang, Y.-W. Dai, and Z.-Q. Wang, "Fast and robust forensics for image region-duplication forgery," *Acta Automatica Sinica*, vol. 35, pp. 1488–1495, 2010.
[18] G. Li, Q. Wu, D. Tu, and S. Sun, "A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2007, pp. 1750–1753.
[19] W. Luo, J. Huang, and G. Qiu, "Robust detection of region-duplication forgery in digital image," in *Proc. Int. Conf. Pattern Recognition*, 2006, pp. 746–749.
[20] B. Mahdian and S. Saic, "Detection of copy-move forgery using a method based on blur moment invariants," *Forensic Sci. Int.*, vol. 171, pp. 180–189, 2007.
[21] V. Christlein, C. Riess, and E. Angelopoulou, "A study on features for the detection of copy-move forgeries," in *Sicherheit 2010*, F. C. Freiling, Ed. Bonn: Gesellschaft für Informatik e.V., 2010, pp. 105–116.
[22] S. Bayram, H. T. Sencar, and N. Memon, "An efficient and robust method for detecting copy-move forgery," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2009, pp. 1053–1056.
[23] W. Li and N. Yu, "Rotation robust detection of copy-move forgery," in *Proc. IEEE Int. Conf. Image Processing*, 2010, pp. 2113–2116.
[24] S. Bravo-Solorio and A. K. Nandi, "Automated detection and localisation of duplicated regions affected by reflection, rotation and scaling in image forensics," *Signal Process.*, vol. 91, no. 8, pp. 1759–1770, 2011.
[25] Q. Wu, S. Wang, and X. Zhang, "Log-polar based scheme for revealing duplicated regions in digital images," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 559–562, Oct. 2011.
[26] J. Wang, G. Liu, H. Li, Y. Dai, and Z. Wang, "Detection of image region duplication forgery using model with circle block," in *Proc. Int. Conf. Multimedia Information Networking and Security*, 2009, vol. 1, pp. 25–29.
[27] G. Liu, J. Wang, S. Lian, and Z. Wang, "A passive image authentication scheme for detecting region-duplication forgery with rotation," *J. Netw. Comput. Applicat.*, vol. 34, no. 5, pp. 1557–1565, 2011.
[28] H. Huang, W. Guo, and Y. Zhang, "Detection of copy-move forgery in digital images using SIFT algorithm," in *Proc. Pacific-Asia Conf. Computational Intelligence and Industrial Applications*, 2008, vol. 2, pp. 272–276.
[29] X. Pan and S. Lyu, "Region duplication detection using image feature matching," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 4, pp. 857–867, Dec. 2010.
[30] I. Amerini, L. Ballan, R. Caldelli, A. D. Bimbo, and G. Serra, "A SIFT-based forensic method for copy-move attack detection and transformation recovery," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 1099–1110, Sep. 2011.
[31] S.-J. Ryu, M.-J. Lee, and H.-K. Lee, "Detection of copy-rotate-move forgery using Zernike moments," in *Information Hiding*, ser. Lecture Notes in Computer Science, R. Böhme, P. Fong, and R. Safavi-Naini, Eds. Berlin, Heidelberg, Germany: Springer-Verlag, 2010, vol. 6387, pp. 51–65.
[32] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," in *Proc. Int. Conf. Very Large Data Bases*, 1999, pp. 518–529.
[33] M. K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. Inf. Theory*, vol. IT-8, no. 2, pp. 179–187, Feb. 1962.
[34] H.-S. Kim and H.-K. Lee, "Invariant image watermark using Zernike moments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 766–775, Aug. 2003.
[35] C.-H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 496–513, Apr. 1988.
[36] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, May 1990.
[37] F. Zernike, "Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode," *Physica*, vol. 1, pp. 689–704, 1934.
[38] Z. Chen and S.-K. Sun, "A Zernike moment phase-based descriptor for local image representation and matching," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 205–219, Jan. 2010.
[39] S. X. Liao and M. Pawlak, "On the accuracy of Zernike moments for image analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1358–1364, Dec. 1998.
[40] A. Andoni, M. Datar, N. Immorlica, P. Indyk, and V. Mirrokni, "Locality-sensitive hashing using stable distributions," in *Nearest-Neighbor Methods in Learning and Vision*, G. Shakhnarovich, T. Darrell, and P. Indyk, Eds. Cambridge, MA, USA: MIT Press, 2005, ch. 3.
[41] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 6, no. 24, pp. 381–395, 1981.
[42] H. Anton and R. C. Busby, *Contemporary Linear Algebra*. Hoboken, NJ, USA: Wiley., 2003.
[43] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1841–1854, Dec. 2012.
[44] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system—The ins and outs of organizing BOSS," in *Information Hiding*, ser. Lecture Notes in Computer Science, T. Filler, T. Pevný, S. Craver, and A. Ker, Eds. Berlin, Heidelberg, USA: Springer-Verlag, 2011, vol. 6958, pp. 59–70.
[45] E2lsh Package [Online]. Available: http://web.mit.edu/andoni/www/LSH/
[46] R. Böhme and M. Kirchner, "Counter-forensics: Attacking image forensics," in *Digital Image Forensics*, H. T. Sencar and N. Memon, Eds. New York, NY, USA: Springer, 2013.
[47] H. C. Nguyen and S. Katzenbeisser, "Security of copy-move forgery detection techniques," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2011, pp. 1864–1867.
[48] 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. Inf. Forensics Security*, vol. 8, no. 4, pp. 593–607, Apr. 2013.
[49] M. Kirchner, "Efficient estimation of CFA pattern configuration in digital camera images," in *Media Forensics and Security II*, ser. Proceedings of SPIE, N. D. Memon, J. Dittmann, A. M. Alattar, and E. J. Delp, Eds. Bellingham, WA, USA: SPIE, 2010, vol. 7541, p. 754111.
[50] T. Bianchi and A. Piva, "Detection of non-aligned double JPEG compression based on integer periodicity maps," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 842–848, Apr. 2012.
[51] E. Kee and H. Farid, "A perceptual metric for photo retouching," in *Proc. National Academy of Sciences*, 2011, vol. 108, no. 50, pp. 19 907–19 912.

Seung-Jin Ryu received the B.S. degree in computer science from the University of Seoul, Seoul, Republic of Korea, in 2007, and has been in the Integrated Master's Ph.D. Program in Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. He is currently pursuing the Ph.D. degree in the Multimedia Computing Laboratory, Department of Computer Science, KAIST. His research interests include multimedia forensics, data hiding, and steganography and steganalysis.

Matthias Kirchner is a member of the IT Security Research Group at the Department for Information Systems, University of Münster, Germany. He received the Ph.D. degree in computer science from the Technical University of Dresden, Germany, in 2011, and he was a postdoctoral research fellow at the International Computer Science Institute in Berkeley, CA. His research interests focus on multimedia security, and in particular on all aspects of multimedia forensics and counter-forensics.

Min-Jeong Lee received the B.S. degree in computer engineering from Kyungpook National University, Republic of Korea, in 2006, and the Ph.D. degree (Integrated Master's Ph.D. Program) in computer science from Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea, in 2011. She is currently working as a senior researcher at SK Telecom. Her research interests are focused on mobile application security, secure browsing, and video watermarking with particular attention to multimedia forensics.

Heung-Kyu Lee received the B.S. degree in electronics engineering from the Seoul National University, Seoul, Korea, in 1978, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology (KAIST), in 1981 and 1984, respectively. Since 1986, he has been a professor in the Department of Computer Science, KAIST. His research team was selected as the Lab. of Excellence (LOE), the National Research Lab. (NRL), and the National Leading Research Lab. (NLRL) by Korea Industry Technology Foundation in 2005 and National Research Foundation of Korea in 2007 and 2012, respectively. His major interests are information hiding and multimedia forensics. Prof. Lee is working with many government institutes and companies to develop the practical products in Korea. Since 2005, he received more than 20 research grants totaling over 6 million U.S. dollars for projects on information hiding and multimedia forensics. He has authored/coauthored over 200 international papers.