https://doi.org/10.2352/ISSN.2470-1173.2019.5.MWSF-531

Dictionary Learning and Sparse Coding for Digital Image Forgery Detection

[a]Mohammed Aloraini [1], [b]Lingdao Sha [1], [c]Mehdi Sharifzadeh [1], [d]Dan Schonfeld [1]
[a]malora2@uic.edu, [b]lsha3@uic.edu, [c]mshari5@uic.edu, [d]dans@uic.edu
[1] Department of Electrical and Computer Engineering, University of Illinois at Chicago, 851 S Morgan St, Chicago, IL 60607, USA

Abstract

Nowadays, digital images are used as critical evidence for judgment, but they can be forged using image processing tools with invisible traces and little effort. Hence, it is very important to determine the authenticity of these digital images. In this paper, we propose a novel approach that uses dictionary learning and sparse coding to detect digital image forgery. We experimented with two popular data sets to determine how effectively and efficiently our approach detects digital image forgery compared to previous approaches. The results show that our approach not only outperforms these approaches in terms of Precision, Recall, and F1 score, but it is also more robust against compression and rotation attacks. Also, our approach detects forgery significantly faster than previous approaches since it uses a sparse representation that dramatically reduces the feature dimensionality by a factor of more than 20.

Introduction

In the past, we had confidence in the integrity of digital images. However, today we are living in an age in which everyone is exposed to abundant imagery. From fashion blogs to scientific journals, image processing tools like Adobe Photoshop and Microsoft Paint are used to facilitate the purpose of being attractive or expressive. Most of the powerful editing tools are user-friendly, but it also causes an increase in digital crimes. As a result, digital imaging tools have put that confidence at risk.

This issue leads to an increasing concern about the originality of digital images contents and the need to develop effective techniques to evaluate the originality, integrity, and authenticity of of these digital images. Over the past decade, several approaches have been proposed to authenticate digital images. These approaches are mainly classified into active and passive detection approaches. In the active approaches, a digital watermark or signature is embedded into digital images in advance. The downsides of these approaches are that embedding decreases the image quality, and its usage is very limited since the digital watermark must be inserted at the time of taking an image. In the passive approaches [1], image statistics are analyzed to validate an image instead of inserting a pre-embedded signature[2].

Passive approaches have been proposed to detect two types of forgery, which are splicing and copy-move forgeries. Image splicing is a process of combining regions from two or more images to form a forged image [3, 4]. The most common forgery is image copy-move forgery (CMF),i.e., a part of the image is copied and pasted in another part of the same image to add or hide objects [5]. The CMF is diffused since a forger uses only one image to make a forgery. Two examples of image copy move forgery are illustrated in Fig.1 where the solid square shows the forgery parts of these images.

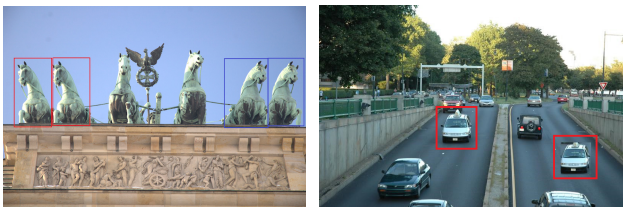


Figure 1: Two examples of image copy-move forgery. Two horses are cloned on the left image and one car is cloned on the right image. Cloned sections are squared in different colors.

Current existing forensic approaches for image copy-move forgery can be divided into two classes: block-based and keypoint-based approaches [6]. First, block-based approaches divide an image into patches, and then detect forgery by looking for the similar patches. The representative approaches are DCT [7], PCA [8], DWT with KPCA [9], and Zernike moment [10]. Second, keypoint-based approaches begin with extracting interest points (keypoints), such as edges and corners from an image and then finding similarities between these points [11]. The representative approaches are SIFT [12, 13] and SURF [14].

Other works combine block-based with keypoint-based approaches to enhance detection results [15, 16, 17]. In [15], Jian Li et al. introduced expectation- maximization (EM) stage after segmenting an image into patches to reduce transform estimation error between copy and original areas. Although this stage improves detection results, it imposes a large computational cost because of the iterative procedure in the EM algorithm. In [17], combining different detector approaches was proposed, along with extensions for behavior knowledge space representation fusion, in order to enhance detection accuracy. However, this approach is computationally expensive since it combines number of detection approaches.

In this paper, problem of detecting image Copy-Move Forgery (CMF) is investigated along with the computational cost, and we propose an approach that is based on a sparse representation of keypoint descriptors to reduce the dimensionality of these descriptors and to remove noisy features from them. We utilize sparse coding,i.e., an unsupervised algorithm aim to learn set of overcomplete basis vectors (atoms) to represent data efficiently. We use sparse coding instead of traditional dimension reduction techniques such as PCA for two main reasons. *First*, sparse cod-

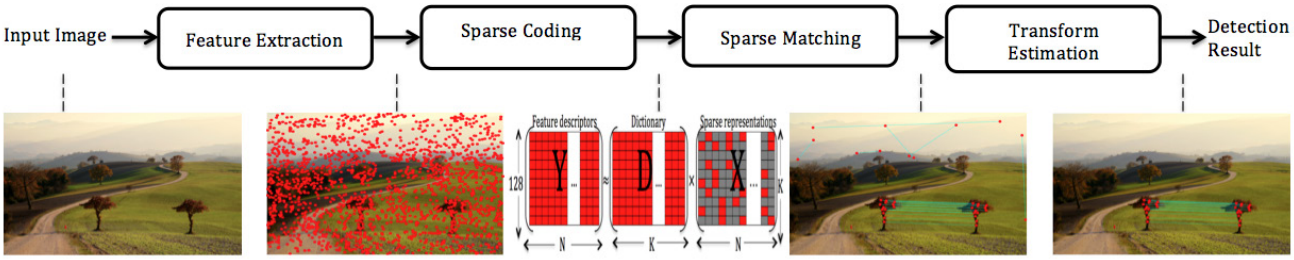


Figure 2: Flowchart of the proposed approach.

ing is able to learn overcomplete atoms and doesn't require these atoms to be orthogonal. *Second*, sparse coding has been widely used in image classification and pattern recognition and it has achieved promising performance. Thus, it is suitable for image forgery problem since it is binary classification, i.e., tampered vs. authentic image.

The contributions of our approach can be summarized as follows:

- We have proposed a novel matching criteria based on dictionary atoms that results in a more effective and efficient forgery detection approach when compared to the original SIFT matching model [12].
- Our approach is robust against compressions and rotations attacks since it uses sparse representations that better fit features descriptors of an image.
- Our approach is scalable since its computational complexity is significantly reduced by using low-dimensional feature vectors of an image.

The rest of the paper is organized as follows: Section 2 presents our sparse representation approach. Section 3 presents our experimental results, followed by conclusion in section 4.

Proposed approach

We briefly describe our approach in the following steps, as illustrated in Fig 2. First, we extract Scale Invariant Feature Transform (SIFT) [18] from an image. Second, K means-Singular Value Decomposition algorithm (K- SVD) [19] is utilized to obtain a sparse representation of SIFT descriptors. Third, the matching process is performed by finding similarities between these sparse features. Next, agglomerative hierarchical clustering [20] is applied on spatial locations of the matched points to identify possible cloned areas. Finally, geometric transformation estimation is obtained between these cloned areas by using RANdom SAmple Consensus algorithm (RANSAC) [21]. An image is forged if a uniform transformation matrix can be obtained between any two matched areas.

Features Extraction

A scale invariant feature transform algorithm (SIFT) [18] is used for keypoint detection and description since SIFT is more robust against scaling, rotating, and illumination. The SIFT algorithm starts with generating 4 octaves (i.e., images with the same size) each of which has 5 images with 5 different blur levels(scales). In each octave, any two consecutive images are subtracted to obtain Difference of Gaussian (DoG) images. Then, maxima and minima (i.e., keypoints) in DoG images are detected by comparing neighboring pixels in the same scale and neighboring scales. If any keypoint has an intensity below a predefined threshold or lies along an edge, it is rejected. Subsequently, a 16×16 window is taken around the keypoint and broken into a 4×4 window. Then, gradient magnitudes and orientations are calculated to generate an 8-bin histogram, which is used to form $128(4 \times 4 \times 8)$ elements as a feature vector,i.e., descriptor.

Sparse Coding

Sparse representation is a method of representing data via a linear combination of dictionary atoms (columns). Given $Y \in \mathscr{R}^{128 \times N}$ as SIFT descriptors of an image,the goal is to find a dictionary with K atoms $D \in \mathscr{R}^{128 \times K}$ and a representation $X \in \mathscr{R}^{K \times N}$ such that $Y \approx DX$ and X is sparse enough (equation 1). We use an adaptive dictionary learning method called K-SVD [19] that has two stages: sparse coding and dictionary update. First, the K-SVD starts with initializing random dictionary D. Then, during sparse coding stage, it finds the best sparse representations X using an orthogonal matching pursuit algorithm [22], given the current dictionary D. Next, during dictionary update stage, it updates dictionary atoms one at a time by using the current sparse representations X. Then, it iterates until the algorithm converges or reaches a predefined number of iterations.

$$\min_{D,X}\{\|Y - DX\|_F^2\} \quad \text{s.t.} \quad \forall i, \|x_i\|_0 \leq S \tag{1}$$

F denotes the Frobenius norm, and $\|\cdot\|_0$ is the $\mathscr{L}^0$ pseudo-norm that counts the non-zero entries.

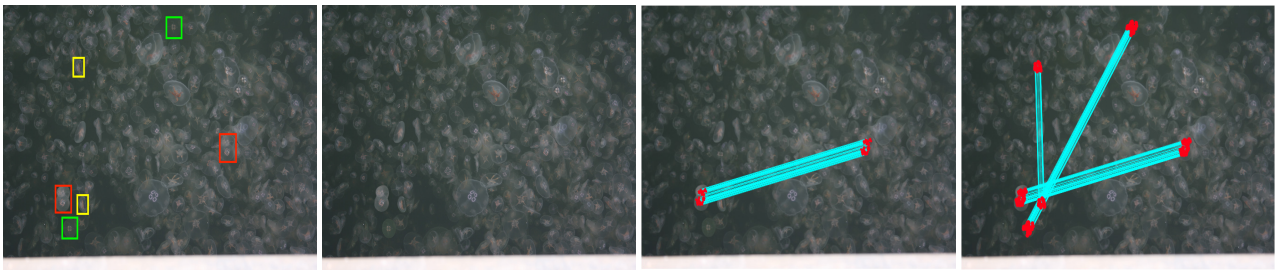
By using the K-SVD algorithm, we approximate SIFT features (128 elements) based on just 6 dictionary atoms

Sparse Matching

We have experimentally observed that similar keypoints tend to use the same dictionary atoms in their sparse representations but with different sparse coefficients. For this reason, we propose a novel matching criteria to detect multiple copies of the same features (keypoints), where a keypoint matches other keypoints if their sparse representations are obtained by using the same dictionary atoms. In other words, if a_i is a vector that locates non zero entries in $x_i \in \mathscr{R}^K$ which is a sparse representation of a keypoint descriptor $y_i \in \mathscr{R}^{128}$ for a keypoint i, then the keypoint i matches another keypoint j if and only if $a_i = a_j$. We obtain the set of matched keypoints by iterating over sparse representations of the keypoints descriptors in an image.

Geometric Transformation Estimation

Given the matched keypoints in an image, we employ agglomerative hierarchical clustering [20] on spatial locations of the keypoints. Hierarchical clustering begins with one keypoint in each cluster, then it combines the closest pair of clusters into a



(a) (b) (c) (d)

Figure 3: Detection results on an image from IMD data set. (a) Forged image with three tampered regions. (b) Detection result of G2NN-SIFT and PCA-SIFT. (c) Detection result of Segmented SIFT. (d) Detection result of the proposed approach.

single cluster, and computes the distances between the new cluster and all the other clusters. The clustering process is repeated until a threshold condition is reached to segregate original regions from copy regions. After clustering is performed, we estimate the affine transformation matrix H between any pair of matched clusters. Let x_i and x_i' be the homogenous coordinates of the matched keypoints in the copy region and original region, respectively. Then the geometric relationship between them is defined as follows:

$$x_i' = Hx_i \tag{2}$$

Considering the existence of outliers (mismatched keypoints), we perform matrix estimation using the RANSAC algorithm [21]. The algorithm estimates the matrix H by randomly selecting three matched pairs, and then it transforms all other points using H. A pair of matched keypoints is an inlier if the distance between the keypoint and the corresponding transformed one is less than a predefined threshold. The process is repeated until a predefined number of iterations is achieved. Finally, the estimated matrix H, which results in a large number of inliers, is elected.

Experimental results

We conduct our experiments on two public data sets. The first data set is the Image Manipulation Data set (IMD) [6], which consists of 48 original images, 48 plain CMF images, and 1392 images that have a single attack, i.e., rotation or noise addition, or JPEG compression. The second data set is MICC-F600 [12], which contains 440 original images and 160 forged images. The 160 forged images consist of 40 images that have one duplicated region, 40 images that have two or three duplicated regions, 40 images that have one duplicated region with 30° rotations, and 40 images that have one duplicated region with 30° rotations and 120% scaling. We have chosen these two data sets because they are used in the validation of a recent work [15] and according to Amerini et al. [12], the MICC-F600 data set is the most challenging data set among the other data sets that were constructed by them.

Evaluation Metric

By defining T_P as the correctly detected forged images, F_P as original images that have been incorrectly detected as forged and F_N as falsely missed forged images, we compute *Precision*, *Recall*, and $F1$ as follows:

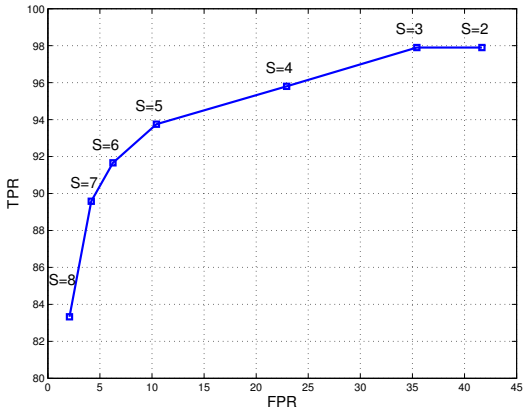


Figure 4: ROC curve: True positive vs. false positive rates for different sparsity parameter settings using a dictionary with 512 atoms.

$$Precision = \frac{T_P}{T_P + F_P} \tag{3}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{4}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Presision + Recall} \tag{5}$$

Precision shows the probability that a detected forgery is truly a forgery, *Recall* indicates the probability that a forged image is detected, and $F1$ score combines precision and recall in a single value.

Comparison of Detection Results

We compare our approach with three different approaches on the two data sets. One recent work, Segmented SIFT [15], which was introduced briefly in the first section, in addition to G2NN-SIFT [12], which handles multiple matches using g2NN, are selected for comparison. We also impliment PCA-SIFT [23] that reduces the dimension of the feature vector to 20 elements. We have chosen these approaches to compare with a recent work, SIFT-based approach, and dimensionality reduction approach. All these approaches including the proposed approach are implemented on a machine with an *Intel Core i7 with 8-GB RAM*. Readers are referred to [24] for more details about our implantation and source code.

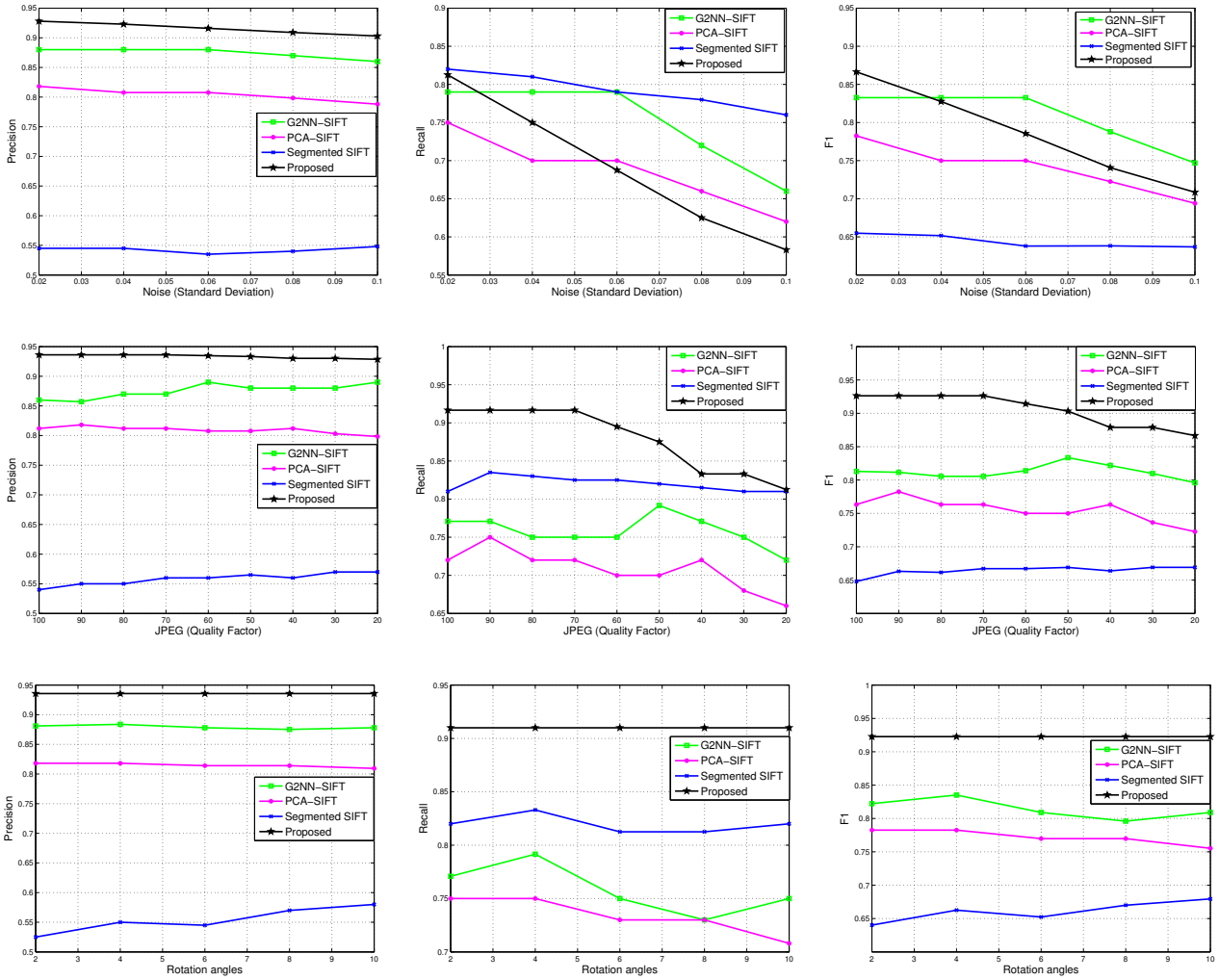


Figure 5: Performance comparison between our approach and the other approaches against three different attacks, which are adding noise, JPEG compression, and rotations. The three columns are corresponding to *Precision*, *Recall*, and $F1$ results, respectively. The three rows are corresponding to adding noise, JPEG compression, and rotations, respectively.

Sparsity Settings

We randomly choose 10% of the datasets to tune the sparsity parameter of our test to achieve the lowest possible feature dimension that leads to high performance. We empirically choose the dictionary with 512 atoms. The receiver operating characteristic curve that is illustrated in Fig.4 suggested that the best tradeoff between the true positive and false positive rates can be achieved when feature size (S) equals 6.

Results on IMD Data Set

First, we evaluate the ability of our approach and the other approaches to detect plain CMF, i.e., a part of the image is copied and pasted in another part of the same image without any attack. An example of detection results on an image is shown in Fig.3. The detection results and average computation times in seconds are shown in Table 1. We observe that our approach not only outperforms the other approaches but also results in low computational time due to its low-dimensional feature vectors, i.e., 6.

Table 1: Results of Plain CMF Detection and Average Computation Times per Image in Seconds on IMD Data Set

Method	Precision (%)	Recall (%)	$F1$	Time (s)
G2NN-SIFT [12]	88.4	79.2	83.5	610
PCA-SIFT [23]	81.8	75.0	78.3	214
Segmented SIFT [15]	70.2	83.3	76.2	719
Proposed	**93.6**	**91.7**	**92.6**	**146**

Next, we evaluate the detection ability of our approach and the other approaches against three different attacks, including noise addition, JPEG compression, and rotation. The experimental results are shown in Fig. 5, which summarizes the detection results for different attacks. We observe that our approach drops linearly when large amounts of noise are added. However, our approach is more robust against JPEG compression and rotation, and it outperforms the other approaches.

Results on MICC-F600 Data Set

We select this dataset to evaluate the detection ability of our approach against combined attacks and large rotation angle,

Table 2: Detection Result on MICC-F600 Data Set

Method	Precision (%)	Recall (%)	$F1$
G2NN-SIFT [12]	84.6	69.0	76.0
PCA-SIFT [23]	83.2	66.3	73.8
Segmented SIFT [15]	86.4	88.1	87.2
Proposed	**94.7**	**94.4**	**94.5**

i.e., 30°. The detection results are shown in Table 2 in temrs of *Precision*, *Recall*, and $F1$, which suggests that our approach achieves superior performance compared to the other approaches. The average computation time is not reported in this table since the sizes of the images in the two datasets are the same.

Conclusion

In this paper, digital image forgery is investigated, and we have proposed a novel approach based on sparse representation of keypoint descriptors to reduce the dimensionality of these descriptors and to remove noisy features from them. Furthermore, we have proposed a new matching criteria that is performed using dictionary atoms instead of ratios between SIFT descriptors. By using this matching criteria, we eliminate efforts of adjusting a threshold. Results show that our approach not only outperforms all the other approaches in terms of *Precision*, *Recall*, and $F1$ score but it is also efficient and more robust against compression and rotation attacks. Our further research will focus on improving our approach to detect images with larger amounts of noise.

References

[1] L. Gaborini, P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Multi-clue image tampering localization," in *Information Forensics and Security (WIFS)*. IEEE, 2014, pp. 125–130.

[2] T.-T. Ng, S.-F. Chang, and Q. Sun, "Blind detection of photomontage using higher order statistics," in *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512)*, vol. 5, May 2004, pp. V–688–V–691 Vol.5.

[3] H. Farid, "Image forgery detection," *IEEE Signal processing magazine*, vol. 26, no. 2, pp. 16–25, 2009.

[4] Z. He, W. Lu, W. Sun, and J. Huang, "Digital image splicing detection based on markov features in dct and dwt domain," *Pattern Recognition*, vol. 45, no. 12, pp. 4292–4299, 2012.

[5] S. Bayram, H. T. Sencar, and N. Memon, "A survey of copy-move forgery detection techniques," in *IEEE Western New York Image Processing Workshop*. IEEE, 2008, pp. 538–542.

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

[7] A. J. Fridrich, B. D. Soukal, and A. J. Lukáš, "Detection of copy-move forgery in digital images," in *in Proceedings of Digital Forensic Research Workshop*. Citeseer, 2003.

[8] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicated image regions," Tech. Rep., 2004.

[9] M. Bashar, K. Noda, N. Ohnishi, and K. Mori, "Exploring duplicated regions in natural images," *IEEE Transactions on Image Processing*, pp. 1–1, 2016.

[10] S.-J. Ryu, M.-J. Lee, and H.-K. Lee, "Detection of copy-rotate-move forgery using zernike moments." in *Information hiding*, vol. 6387. Springer, 2010, pp. 51–65.

[11] A. Zandi, Mohsen Aznaveh and A. Talebpour, "Iterative copy-move forgery detection based on a new interest point detector," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, 2016.

[12] 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 Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1099–1110, Sept 2011.

[13] H. Huang, W. Guo, and Y. Zhang, "Detection of copy-move forgery in digital images using sift algorithm," in *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, vol. 2, Dec 2008, pp. 272–276.

[14] B. L. Shivakumar, L. Dr, and S. S. Baboo, "Detection of region duplication forgery in digital images using surf," *International Journal of Computer Science Issues*, 2011.

[15] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 507–518, March 2015.

[16] E. Silva, T. Carvalho, A. Ferreira, and A. Rocha, "Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes," *Journal of Visual Communication and Image Representation*, vol. 29, pp. 16–32, 2015.

[17] A. Ferreira, S. C. Felipussi, C. Alfaro, P. Fonseca, J. E. Vargas-Muñoz, J. A. dos Santos, and A. Rocha, "Behavior knowledge space-based fusion for copy–move forgery detection," *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4729–4742, 2016.

[18] D. G. Lowe, "Object recognition from local scale-invariant features," p. 11501157, 1999.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: Design of dictionaries for sparse representation," in *IN: PROCEEDINGS OF SPARS05*, 2005.

[20] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning – data mining, inference, and prediction," 2003.

[21] M. A. Fischler and R. C. Bolles, "Random sample consensus," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[22] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *Information Theory, IEEE Transactions on*, p. 4688, 2011.

[23] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II–II.

[24] M. A. et al., "Source code of sparse representation approach and experimental results," https://www.dropbox.com/s/wigxvhvv5iyapn7/ImageForgeryDetection.zip?dl=0, May 2018.

Author Biography

Mohammed Aloraini *received his BS in electrical engineering from Qassim University in 2011 and the M.S. degree in electrical and computer engineering from University of Illinois at Chicago in 2014. He is now pursuing his PhD in electrical and computer engineering at University of Illinois at Chicago. His current research interests include multimedia forensics and information security.*

Lingdao Sha *received his BS in electrical and computer engineering from Beijing University of Posts and Telecommunications in 2011 and the Ph.D. degree in electrical and computer engineering from University of Illinois at Chicago in 2018. His research interests are image processing, medical image processing and recognition, 3-D images, sparse coding and deep learning.*

Mehdi Sharifzadeh *received his BS in electrical engineering from Sharif University of Technology in 2012. Currently, he is a PhD student and researcher in electrical and computer engineering at University of Illinois at Chicago. His current researches are in machine learning, and problems in image processing and computer Vision.*

Dan Schonfeld *received the B.S. degree in electrical engineering and computer science from the University of California at Berkeley in 1986 and the M.S. and Ph.D. degrees in electrical and computer engineering from the Johns Hopkins University, Baltimore, MD, in 1988 and 1990, respectively. In 1990, he joined the University of Illinois at Chicago, where he is currently a Professor in the Department of Electrical and Computer Engineering. He has authored over 120 technical papers in various journals and conferences. His current research interests are in multi-dimensional signal processing, image and video analysis, computer vision, and genomic signal processing.*

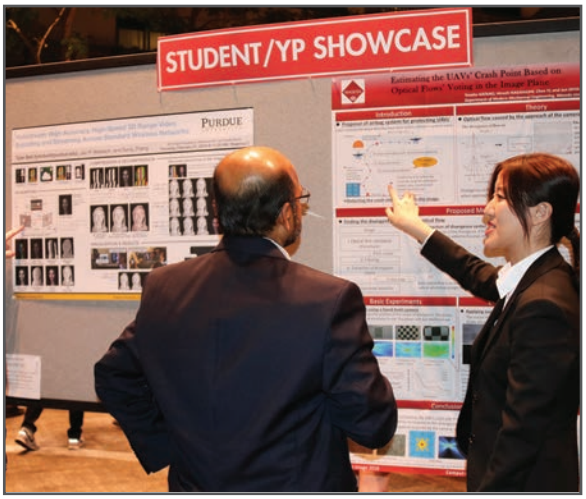
JOIN US AT THE NEXT EI!

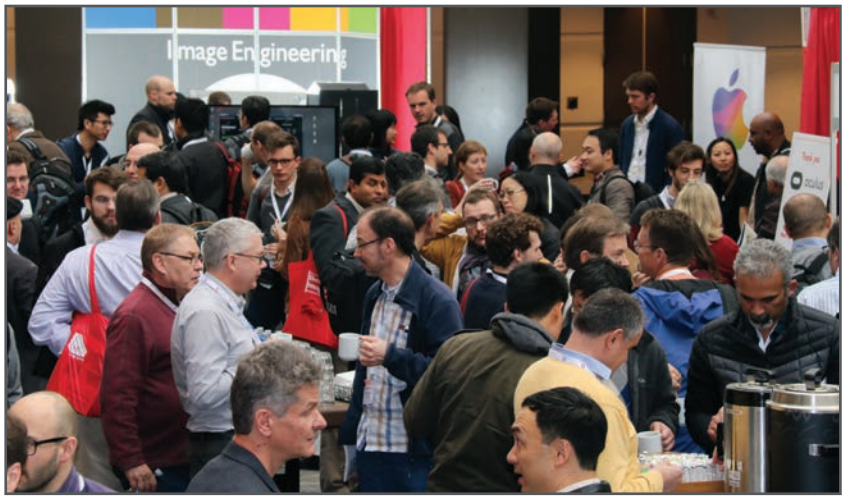
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org