ROBUST IMAGE RECAPTURE DETECTION USING A K-SVD LEARNING APPROACH TO TRAIN DICTIONARIES OF EDGE PROFILES

Thirapiroon Thongkamwitoon, Hani Muammar, and Pier Luigi Dragotti

Electrical and Electronic Engineering Department, Imperial College London, United Kingdom

ABSTRACT

A professionally recaptured image from an LCD monitor can be, visually, very difficult to distinguish from its original counterpart. In this paper we show that it is possible to detect a recaptured image from the unique nature of the edge profiles present in the image. We leverage the fact that the edge profiles of single and recaptured images are markedly different and we train two alternative dictionaries using the K-SVD approach. One dictionary is trained to provide a sparse representation of single captured edges and a second for recaptured edges. Using these two learned dictionaries, we can determine whether a query image has been recaptured. We achieve this by observing the type of dictionary that gives the smallest error in a sparse representation of the edges of the query image. Experiments conducted show that the proposed algorithm is capable of detecting recaptured images with a high level of accuracy and copes well with a wide range of natural images.

Index Terms— Recapture Detection, Image Forensics, Acquisition Chains, K-SVD, Edge Profiles, Blurring Model

1. INTRODUCTION

Most images, today are stored in a digital format due to advantages in transmission, compression and ease of duplication. Digital images that are printed or displayed on a monitor are sometimes recaptured and saved digitally. For example, a print may be scanned using a flat-bed scanner or an image may be recaptured from an LCD monitor with a mobile phone camera. One reason that an image may be recaptured is that the original digital version may not be available. If an image is recaptured professionally using high quality photographic equipment, the resulting recaptured image may be very difficult to differentiate from the original version [1]. Footprints, such as noise patterns or JPEG artefacts, left either by the acquisition process or by post-processing operations, are vital to many digital forgery detection algorithms and may be modified or eliminated by the recapture process.

Many image recapture detection techniques have been proposed in the literature, recently. Yu et al. [2] and Yin and Fang [3] have addressed the detection of recaptured images from printed material using specularity and dithering effects, respectively. Cao and Cot [1] describe a method for detecting recaptured images from LCD monitors using multiple features including colour saturation, loss of detail, and fine textures added by the recapture process. Ke et al. [4] use 136 features to train a support vector machine to detect recaptured images taken from an LCD screen. They test their algorithm on a dataset of images recaptured using smart-phone cameras [5]. Yin et al. [6] describe a detector for images recaptured from LCD screens that analyses noise features and applies the MBFDF algorithm to detect double JPEG compression. If an image recaptured from an LCD monitor contains aliasing patterns, recapture detection can be performed by searching for peaks in the DFT of the recaptured image noise residual [7, 8].

The use of edge profiles to detect recapture has been explored theoretically in [9] and a practical algorithm developed [10]. A limitation of the algorithm in [10] was that it required the manual selection of regions in the image that contained suitable edges. This meant that the detection performance of the algorithm depended on the regions of interest selected by the user. Furthermore, the algorithm required training and test images that contained well defined, sharp, edges. In this paper, we propose an algorithm for image recapture detection using the K-SVD dictionary learning approach. Sets of atoms are built based on dictionary training using features from real images. The robustness of the algorithm is improved relative to previous work [10] since our proposed method is applicable to most random natural images. Edge features are extracted and selected without human supervision thereby making the algorithm completely automatic. For the completeness of the literature review it is worth mentioning that the area of recaptured video detection (e.g. [11, 12, 13, 14]) has, in recent years, received considerable attention.

The paper is organized as follows: Section 2 explains how edges become blurred during the acquisition process and how the distortion is modelled mathematically. Next, our proposed technique is presented in Section 3. The experiments on dictionary training and performance evaluation are discussed in Section 4 before we conclude the paper in Section 5.
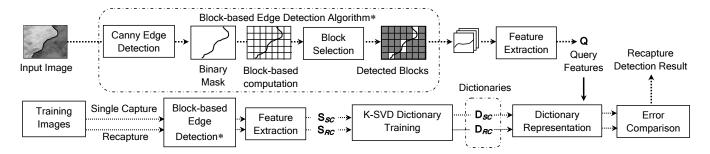
Fig. 2. Diagram of the proposed recapture detection algorithm. Note that the block-based edge detection algorithm (*) used for the input image and training images is the same.
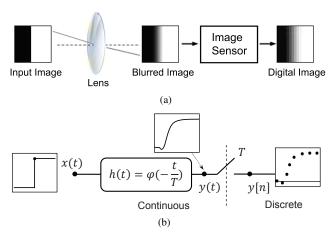

Fig. 1. The modelling of blurriness represented by (a) a simplified acquisition process and (b) a corresponding signal diagram illustrating the A/D conversion process.

2. BLURRINESS MODEL

In our model the observed straight edge $x(t)$ is projected on an image sensor by a camera lens that has been adjusted to provide an optimally focused image as shown in Figure 1(a). Because the edge can be fully described by its cross sectional profile we can simplify the description of an edge feature to a one-dimensional representation. A degree of blur is introduced into the edge due to lens imperfections, such as spherical aberration. If the lens aperture setting is very small, blurring due to diffraction may also be introduced. In-camera post processing such as CFA demosaicing, noise reduction, sharpening and contrast enhancement may introduce additional distortion into the edge. The blurriness patterns, therefore, are unique and are determined by the intrinsic properties of the image acquisition devices. The measurement techniques to model blurriness introduced to images have been proposed in the literature (e.g. [15, 16, 17, 18]). We model the blurriness introduced in a perfect edge using the filter function $h(t)$ as shown in Figure 1(b). The distorted edge is given by $y(t) = x(t) * h(t) = x(t) * \varphi(\frac{t}{T} - n)$ where $\varphi(t)$ is a reversed version of the impulse response $h(t)$. The blurred edge $y(t)$ is then sampled with a uniform sampling period T in order to obtain a discretized edge $y[n] = \langle x(t), \varphi(t/T - n) \rangle$. This suggests that when the input $x(t)$ is acquired by different devices, the differences between outputs, $y[n]$, are determined by the uniqueness of the corresponding φ of each device.

The recapture process is described by a cascade of acquisition, reproduction, and acquisition processes that can be represented by a series of A/D, D/A, and A/D conversions respectively. The reproduction process converts discrete signals back to the continuous domain. During the D/A conversion, a perfect construction generally cannot be achieved and the edge becomes less sharp. As a result, the reproduced edge $\hat{x}(t) \neq x(t)$. When we recapture $\hat{x}(t)$ through a reacquisition process, the recaptured edge, $\hat{y}[n]$, exhibits different blurring characteristics compared to $y[n]$. We use the increasing degree of blur and the different patterns of distortion to detect image recapture.

3. THE PROPOSED TECHNIQUE

This section presents our proposed framework for recapture detection using a dictionary representation of edge profiles. A working diagram of our proposed algorithm is illustrated in Figure 2. The algorithm comprises the three following important parts:

3.1. Representation of Edge Profiles using a Dictionary

Given a query image, blocks containing edges are identified using an automatic edge detection algorithm (the reader is referred to Figure 2 and the implementation details in Section 4.1). Let $\mathbf{Y} \in \mathbb{R}^{W \times W}$ be a matrix that represents the grayscale values of a block with size $W \times W$. We assume that an edge runs across the block in a horizontal or near horizontal direction. Otherwise, if the edge runs from top to bottom, the block is rotated by $90°$. As a result we have the matrix $\mathbf{Y} = [\mathbf{y_1 y_2} \cdots \mathbf{y_i} \cdots \mathbf{y_W}]$ where $\mathbf{y_i}$ represents a column vector of edge profile at column index i with length W. We define a line spread profile, $\mathbf{q}_i$, as $\mathbf{q}_i = \mathbf{y}_i^{(1)}/\|\mathbf{y}_i^{(1)}\|_2$ where $\mathbf{y}_i^{(1)}$ is the first derivative of $\mathbf{y}_i$. The differentiation is performed in order to suppress the DC bias in the edge profile.

We now assume that the dictionary used is preconstructed. The matrix $\mathbf{D} \in \mathbb{R}^{(W-1) \times K}$ represents an overcomplete dictionary that was generated from known dictionary atoms $\{\mathbf{d}_j\}_{j=1}^K$. Our goal is to find a set of atoms that allow the best approximation of the feature $\mathbf{q}_i \in \mathbb{R}^{W-1}$. We solve the following optimization problem:

$$\min_{x_i} \|\mathbf{q}_i - \mathbf{D}\mathbf{x}_i\| \text{ subject to } \forall i, \|\mathbf{x}_i\|_0 \leq T_0, \quad (1)$$

where the vector $\mathbf{x}_i \in \mathbb{R}^K$ is a sparse matrix that contains the representation coefficients of $\mathbf{q}_i$ and T_0 is the maximum number of atoms permitted. In Figure 3 an example of a query line spread profile and its best linear approximation using three atoms is given.
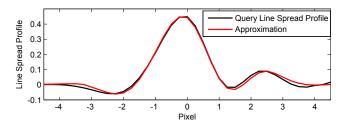

Fig. 3. Representation of a blurring feature (black) using a linear combination (red) of three dictionary atoms.

From the figure, it can be observed that three atoms are sufficient to provide a good approximation to the profile. In optimal sparse representation a dictionary is designed to contain atoms that fit all possible unique feature blurring patterns. One method by which this can be achieved is to build a dictionary by learning from line spread profiles that have been extracted from a sufficiently large set of natural images containing scenes that are representative of everyday life.

3.2. Dictionary Learning

Given that the matrix $\mathbf{S} \in \mathbb{R}^{(W-1) \times N}$ is a collection of all known features from a training set, we build $\mathbf{S}$ by concatenating the N column vectors $\mathbf{q}_i \in \mathbb{R}^{W-1}$ that represent all line spread profiles extracted from the training images. Our objective now is to design the dictionary $\mathbf{D} \in \mathbb{R}^{(W-1) \times K}$ that can best represent all the profiles in the training matrix $\mathbf{S}$. That is:

$$\min_{\mathbf{D}} \|\mathbf{S} - \mathbf{D}\mathbf{X}\|_F^2 \text{ subject to } \forall i, \|\mathbf{x}_i\|_0 \leq T_0, \quad (2)$$

where $\mathbf{X} \in \mathbb{R}^{K \times N}$ is built from the coefficient matrices $\mathbf{x}_i$ used represent the feature $\mathbf{q}_i$ and $i = 1, 2, \ldots, N$.

The learning technique used in this work is the K-SVD dictionary learning algorithm [19]. It is an iterative algorithm that involves two important steps in each iteration: sparse coding and dictionary update. Sparse coding is a method that finds the best T_0 atoms that can represent a signal with minimum error as referred to in Equation (1). We use the orthogonal matching pursuit (OMP) algorithm [20] in order to obtain
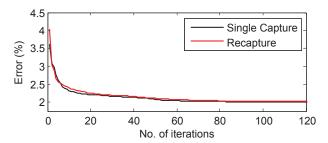

Fig. 4. Representation errors over 120 iterations of K-SVD training using features from single captured and recaptured images

near-optimal sparse coding in our experiment. A dictionary update is then applied. The K-SVD method updates the dictionary atoms one column at a time. It computes the residual error when the atom at column $k : k = 1, 2, \ldots, K$ is deselected. A singular vector which minimizes the residual error is then computed using singular value decomposition (SVD) with the linear least-squared norm properties. This vector is then used to update the kth atom and it is guaranteed that the approximation error can be minimized over each update. The new set of atoms is used for the next computation and the representation error will decrease over a number of iterations.

The training setup is described in Section 4. The initial dictionary used is constructed from a set of line spread functions (LSF) extracted from test targets captured with known devices as presented in our preliminary work [10]. Figure 4 shows the training errors for dictionary learning using line spread profiles extracted from our large set of natural images. The errors for both the single capture and recapture categories can be seen to drop over 120 iterations. This is due to the fact that the atoms, initially created from edge LSFs of known devices using a slanted edge test target, are replaced by the LSFs of edges that are more typically present in natural images. The linear approximations formed from the learned atoms adapt to the edges found in the natural images and, therefore, provide a more accurate representation compared to the atoms formed from input edges that are perfectly sharp.

3.3. Recapture Detection Algorithm

We exploit the fact that, after training, the dictionary elements are adapted to the blurring patterns of the specific training sets. We trained two dictionaries $\mathbf{D}_{SC}$ and $\mathbf{D}_{RC}$ using features $\mathbf{S}_{SC}$ and $\mathbf{S}_{RC}$ extracted from the single captured and recaptured images respectively. Each dictionary is considered to provide an optimal representation of the profiles extracted from edges found in each set of training images, respectively. We now assume that a query image containing edges is available. The line spread profiles are extracted using the algorithm discussed in Section 3.1. Given the matrix $\mathbf{Q} \in \mathbb{R}^{W-1 \times N}$ which represents all line spread profiles extracted from the detected blocks, the decision for recapture

detection can be based on the class of dictionary that gives the smallest representation error. We define $\mathbf{X}_1$ and $\mathbf{X}_2$ as the coefficient matrices obtained from the composition of query feature matrix $\mathbf{Q}$ using the dictionaries $\mathbf{D}_{RC}$ and $\mathbf{D}_{SC}$ respectively. The query image is classified into a recapture group if

$$\|\mathbf{Q} - \mathbf{D}_{RC}\mathbf{X}_1\|_F^2 < \|\mathbf{Q} - \mathbf{D}_{SC}\mathbf{X}_2\|_F^2. \tag{3}$$

Otherwise, the query image is classified to the single capture group.

4. EXPERIMENTS AND RESULTS

Our dataset comprised 9 sets of random single capture images that were respectively captured using a Nikon D40, Nikon D70s, Canon EOS 600D, Olympus E-PM2, Sony RX100, Panasonic Lumix TZ7, Kodak V550 (black body), Kodak V550 (silver body) and a Kodak V610 digital camera. All the images were taken from random naturally occurring scenes with no restrictions on types of edges in the pictures.

The single captured images were recaptured from an LCD monitor using a technique that was derived from a method for recapturing alias-free images[7]. The set of 8 cameras used to perform the recapture were a Nikon D70S, Canon EOS600D, Olympus E-PM2, Sony RX100, Nikon D3200, Canon A700, Canon EOS60D and Panasonic Lumix TZ10. Each set of single captured images was recaptured by each recapture camera resulting in 72 different combinations of recapture. The single capture and recaptured images were divided into training and testing sets. Figure 5 provides an example of the images used in the experiments. Figure 5(a) is a single image captured by a Nikon D70s camera and Figure 5(b) shows its recaptured version taken using a Canon PowerShot A700.
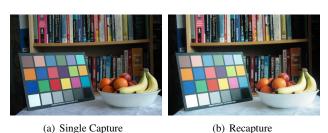

(a) Single Capture   (b) Recapture

Fig. 5. Examples of images used in the experiments (a) Single capture (b) Recapture

4.1. Dictionary Training

A total of 135 single captured and 243 recaptured images were used in the training set-up. A binary map of edges was produced for each image using the Canny edge detector [21]. Blocks of size 16×16 pixels were selected based on the criteria: 1) when sampling the block in the horizontal (vertical) dimension, the block must contain a region that is spanned by a single edge only, although the block may also include regions spanned by two or more edges 2) the region of the block

spanned by a single edge must form 60% or more of the block height (width). This resulted in 79,440 and 117,352 training features for the single capture and recapture groups, respectively. The orthogonal matching pursuit with $T_0 = 3$ atoms was used for sparse coding. The number of atoms chosen provided an optimal balance between the accuracy of query profile approximation and the ability to discriminate between line spread profiles from single and recaptured images.

4.2. Performance Evaluation

A total of 1620 test images were used to evaluate the performance of the proposed recapture detection algorithm. The set comprised 180 single captured images taken with the 9 cameras, and 1440 recaptured images taken with the 8 recapture cameras, where each recapture camera was used to capture each set of single captured images. The results from our performance evaluation are shown in Table 1. A total of 1385 images out of 1440 recaptured images were correctly classified resulting in a true positive rate (TP) of 96.18% and a false negative rate (FN) of 3.82% . A true negative (TN) rate of 92.22% and a false positive (FP) rate of 7.78%, was obtained from the proposed algorithm. The algorithm is, therefore, robust to a variety of scenes and does place specific requirements regarding image content. Our technique is also capable of extracting features using a block based edge detection algorithm that is completely automatic. The method eliminates possible operator bias during edge selection and is, therefore, suitable to large image data sets.

Table 1. Results of the performance evaluation conducted on single and recaptured images from the image test set.

Image Test Set	Classification Results	
	Recaptured	Single Captured
Recaptured	96.18%	3.82%
Single Captured	7.78%	92.22 %

5. CONCLUSIONS

In this paper we have proposed a recapture detection algorithm based on the specificity of sparse representation of features from edges using overcomplete dictionaries. We have shown that the result of detection is given by the class of trained dictionary that provides optimal approximation of query blurring patterns. The experimental results indicate that the algorithm performs well when tested on large sets of natural images. The performance of the proposed method is comparable with earlier work and does not require human supervision or the presence of sharp edges in query images. The algorithm is therefore applicable to a wider variety of natural images.

6. REFERENCES

[1] H. Cao and A.C. Kot, "Identification of recaptured photographs on LCD screens," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 1790–1793.

[2] H. Yu, T.T. Ng, and Q. Sun, "Recaptured photo detection using specularity distribution," in *IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 3140–3.

[3] J. Yin and Y. Fang, "Markov-based image forensics for photographic copying from printed picture," in *Proceedings of the ACM International Conference on Multimedia (MM)*, Nara, Japan, 2012, pp. 1113 – 1116.

[4] Y. Ke, Q. Shan, F. Qin, and W. Min, "Image recapture detection using multiple features," *International Journal of Multimedia and Ubiquitous Engineering*, 2013.

[5] X. Gao, B. Qiu, J.J. Shen, T.T. Ng, and Y.Q. Shi, "A smart phone image database for single image recapture detection," in *Proceeding of the International conference on Digital Forensics and Watermarking (IWDW)*, 2010, pp. 90–104.

[6] J. Yin and Y. Fang, "Digital image forensics for photographic copying," in *Proceedings of the SPIE - The International Society for Optical Engineering*, 2012, vol. 8303, p. 83030F (7 pp.).

[7] H. Muammar and P.L. Dragotti, "An investigation into aliasing in images recaptured from an LCD monitor using a digital camera," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada*, 2013, pp. 2242 – 2246.

[8] A. da Silva Pinto, H. Pedrini, W. Schwartz, and A. Rocha, "Video-based face spoofing detection through visual rhythm analysis," in *The 25th Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2012, pp. 221–228.

[9] T. Thongkamwitoon, H. Muammar, and P.-L. Dragotti, "Reverse engineering of signal acquisition chains using the theory of sampling signals with finite rate of innovation," in *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 2013, pp. 5910 – 5914.

[10] T. Thongkamwitoon, H. Muammar, and P.L. Dragotti, "Identification of image acquisition chains using a dictionary of edge profiles," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2012, pp. 1757–1761.

[11] W. Wang and H. Farid, "Detecting re-projected video.," in *Information Hiding*. 2008, vol. 5284 of *Lecture Notes in Computer Science*, pp. 72–86, Springer.

[12] D.J. Jung, D.K. Hyun, S.J. Ryu, J.W. Lee, H.Y. Lee, and H.K. Lee, ," in *Proceeding of the 10 International conference on Digital Forensics and Watermarking (IWDW)*. 2011, pp. 281–291, Springer.

[13] M. Visentini-Scarzanella and P.L. Dragotti, "Video jitter analysis for automatic bootleg detection," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2012, pp. 101 – 106.

[14] P. Bestagini, M. Visentini-Scarzanella, M. Tagliasacchi, P.L. Dragotti, and S. Tubaro, "Video recapture detection based on ghosting artifact analysis," in *IEEE International Conference on Image Processing (ICIP)*, 2013.

[15] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric.," in *IEEE International Conference on Image Processing (ICIP)*, 2002, pp. 57–60.

[16] J. Dijk, M. van Ginkel, R. J. van Asselt, L. J. van Vliet, and P. W. Verbeek, "A new sharpness measure based on gaussian lines and edges.," 2003, vol. 2756 of *Lecture Notes in Computer Science*, pp. 149–156, Springer.

[17] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb)," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 717–728, 2009.

[18] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (cpbd).," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2678–2683, 2011.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," in *IEEE Transactions on Signal Processing*, Nov. 2006, vol. 54, pp. 4311–4322.

[20] Y. C. Pati, R. Rezaiifar, Y. C. Pati R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27 th Annual Asilomar Conference on Signals, Systems, and Computers*, 1993, pp. 40–44.

[21] J. Canny, "A computational approach to edge detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, vol. 8, pp. 679–698.