A Simple and Effective Image-Statistics-Based Approach to Detecting Recaptured Images from LCD Screens

Kai Wang

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

Abstract

It is now extremely easy to recapture high-resolution and high-quality images from LCD (Liquid Crystal Display) screens. Recaptured image detection is an important digital forensic problem, as image recapture is often involved in the creation of a fake image in an attempt to increase its visual plausibility. State-of-the-art image recapture forensic methods make use of strong prior knowledge about the recapturing process and are based on either the combination of a group of ad-hoc features or a specific and somehow complicated dictionary learning procedure. By contrast, we propose a conceptually simple yet effective method for recaptured image detection which is built upon simple image statistics and a very loose assumption about the recapturing process. The adopted features are pixel-wise correlation coefficients in image differential domains. Experimental results on two large databases of high-resolution, high-quality recaptured images and comparisons with existing methods demonstrate the forensic accuracy and the computational efficiency of the proposed method.

Keywords: Digital image forensics, recaptured image detection, LCD screen, image statistics, correlation coefficient, support vector machine.

*Corresponding author

Email address: kai.wang@gipsa-lab.grenoble-inp.fr (Kai Wang)

1. Introduction

With the increasing popularity and quality of customer digital cameras (especially those equipped on smart phones) and the consistently improved sophistication of image editing software tools, capturing and editing high-resolution and high-quality digital images has become an easy task. While photo editing can be used to further improve the visual or aesthetic quality, it can also be intentionally used for creating falsified images. In general, falsified images do not reflect what happens in the reality and may have big negative impact on the society. For example, a suspect can fabricate a doctored image, attempting to disclose an alibi defence. Another example is that during an election campaign a candidate may anonymously spread an image of his/her opponent showing up in a disadvantageous but fake scenario, so as to ruin the opponent's reputation. Therefore, nowadays people often have doubt on the credibility of the content described by digital photographs. Accordingly, in order to expose doctored images and recover people's trust on authenticity of digital images, the research on *digital image forensics* has received more and more attention during the last decade. Image forensic techniques attempt to computationally assess the authenticity and/or find the origin of a given digital image based on trustworthy scientific methods and appropriate mathematical tools, see (Stamm et al., 2013; Piva, 2013) for two surveys on this quickly evolving topic.

One particular image forensic problem is to distinguish between *single captured* images (*i.e.*, real-world scenes captured by a digital camera) and *recaptured* images from LCD screens (*i.e.*, a single captured image, potentially after being tampered with, is displayed on an LCD screen and then recaptured by a digital camera). This forensic problem has high utility in practical image forensic scenarios. The fact that a given image is a recaptured one is on its own very suspicious. Serious questions arise concerning the motivation of recapturing an image displayed on an LCD screen and even about the authenticity of the displayed image. Indeed, as discussed below and pointed out by other researchers (Cao and Kot, 2010; Thongkamwitoon et al., 2015), image recapturing is often

involved in the creation of a fake image.

When fabricating a fake image, the creator usually seeks a high visual plausibility, *i.e.*, the fake image should look like a genuine one and be able to pass the human visual inspection. Although with current photo editing software tools people can easily create a visually convincing fake image, it is not uncommon to further increase its plausibility by intentionally displaying the tampered image on an LCD screen and then recapturing the displayed image (Cao and Kot, 2010; Thongkamwitoon et al., 2015). There are two main motivations for this additional step of image recapturing. First, the final fake image will be in the form of a "real" photo taken by a digital camera, instead of a potentially edited image saved by using photo editing software. This will make the origin of the fake image look more convincing and also has a very big chance to deceive any forensic detectors that examine the traces of photo editing software left in the image's metadata or in the image itself. Second, the recapturing process is often capable of increasing the visual plausibility of a fake image. It appears that the combined effect of LCD display and camera recapturing would make the image visually smooth and quite natural. This, for instance, can be useful to hide the sharp and unnatural transition at the border of two spliced subimages (Cao and Kot, 2010). The top row of Fig. 1 shows three recaptured fake images which are visually very convincing. However, reliable image recapture forensic methods, such as the one proposed in this manuscript, can assert that these are actually recaptured images from LCD screens whose authenticity is very questionable.

A number of image recapture detection methods have been proposed in the literature. Nearly all of the existing methods are based on strong assumptions about the recapture process, *e.g.*, the recapturing introduces aliasing-like distortions, changes image's color and edge sharpness, or even further increases lossy compression artifacts. By contrast, in this manuscript, we make a very loose assumption about the consequence of recapturing high-resolution and high-quality images from LCD screens and rely on a simple yet discriminative image statistics feature. More precisely, we assume that the recapturing process introduces subtle but detectable statistical difference to a digital image, reflected and captured

Figure 1: In the top row are three recaptured tampered images from the ROSE database (Cao and Kot, 2010; Cao, 2010), which are visually very plausible and convincing. The corresponding genuine image are shown in the bottom row. In the three tampered images (from left to right), color alteration, object removal and object insertion have been respectively used for creating the fake image.

by features consisting of simple pixel-wise correlation coefficients (CC) in image differential domains. We validate the proposed method on two large databases comprising high-resolution and high-quality recaptured images, and one of the two databases includes recaptured tampered images of high visual plausibility (some examples are shown in the top row of Fig. 1). Experimental results and comparisons with state-of-the-art algorithms demonstrate the forensic accuracy and the computational efficiency of our method.

The remainder of this manuscript is organized as follows. Section 2 reviews existing methods for detecting recaptured images from LCD screens and the common data sets for testing and evaluating such methods. In Section 3, we present the technical details of the proposed forensic method, in particular the design of the adopted image statistics feature. Experimental results are reported in Section 4. We conclude the manuscript and suggest several future working directions in Section 5.

2. Related Work

Nearly all the review of prior work as given in relevant published papers provides an overview of this research field in a method-by-method manner. Here we summarize existing methods in a different way, classifying them as those utilizing strong prior knowledge about the alterations that may exist in a recaptured image and those making a weak assumption about the consequence of image recapturing. At the end of this section, we will also mention the main data sets on which researchers test and compare their methods for recaptured image detection.

2.1. Methods based on detection of specific alterations

The first group of methods is based on strong prior knowledge about the alterations that can be found in a recaptured image when compared to a single captured one. It is worth mentioning that despite the computational difference between a pair of single captured and recaptured images, according to two subjective studies reported by Cao and Kot (2010) and Mahdian et al. (2015), human observers are not good at recognizing recaptured images in a single-stimulus experimental setting, *i.e.*, during subjective studies images are shown one-by-one, as in real-world scenarios, but not in pair of corresponding single captured and recaptured images.

One of the most important alterations is the *blurriness*. In general, the recaptured image looks smoother than the corresponding single captured image. This is in part due to the limited spatial resolution of the LCD screen (*i.e.*, content of a recaptured image), in contrast to the intrinsic continuous nature of the physical world (*i.e.*, content of a single captured image). However, this blurriness remains rather natural and largely unnoticeable for human observers especially in a single-stimulus setting, as shown in the subjective studies mentioned above. Researchers have designed specific and distinctive features for detecting blurriness alteration. Popular features include descriptors of high-frequency coefficients in a transformed domain (Cao and Kot, 2010), a blind

image smoothness measure borrowed from image quality assessment literature (Gao et al., 2010), histogram of image local difference (Ke et al., 2013), and a feature combining edge spread width and the approximation error difference of edge profiles under two sparse dictionaries learned from training samples of single captured and recaptured images (Thongkamwitoon et al., 2015).

Another important alteration is the *aliasing-like distortion*. It is not uncommon to introduce aliasing-like patterns into a recaptured image, especially when the recapturing parameters are not carefully controlled. The main reasons for this aliasing-like alteration include the periodicity of LCD cells, some relevant properties such as the LCD polarity inversion and the periodic recharging, as well as the interplay between the periodicity of LCD cells and that of the image sensor (Cao and Kot, 2010; Thongkamwitoon et al., 2015). One of the popular features used to detect aliasing-like alterations is the well-known Local Binary Pattern (LBP) descriptor originally propose by Ojala et al. (2002) for texture classification. It has been demonstrated in (Cao and Kot, 2010) that LBP is capable of capturing the subtle difference in fine local textures between single captured and recaptured images, and this difference is mainly due to the aliasing-like patterns. Following this observation, Ke et al. (2013) make use of a variant of LBP, the so-called Center-Symmetric Local Binary Pattern (CS-LBP) (Heikkilä et al., 2009), for recaptured image detection. The aliasing-like alteration can also be detected by using the cyclostationarity theory for which the basic idea is to check whether a signal has a high correlation with one of its translated versions in the spectral domain (Mahdian et al., 2015), or by extracting discriminative features after applying a specific aliasing enhancement algorithm (Li et al., 2015).

According to Thongkamwitoon et al. (2015), blurriness and aliasing are the two most important alterations that remain least dependent on the image content. In the same paper, the authors also show that if the recapturing parameters are properly chosen, the aliasing distortion can be largely decreased to a nearly invisible level. This is actually the main motivation of their image recapture detection method based on an edge sharpness feature to assess blurriness.

Other alterations can be introduced into a recaptured image. The first is *color* alteration, which is mainly due to the specific and often limited color range of LCD screens and to the different ambient lighting conditions when taking the single captured and recaptured images. For recaptured image detection, different color descriptors have been considered, including color histograms (Gao et al., 2010), color moments (Cao and Kot, 2010; Ke et al., 2013; Ni et al., 2015), color energy ratios (Cao and Kot, 2010), and chromaticity covariance matrix (Gao et al., 2010). Second, researchers also make assumptions that single captured and recaptured images have different *noise* characteristics. Here the noise mainly consists of display and camera noises, but in a general sense also includes the aforementioned aliasing-like distortion. Noise features for recaptured image detection are typically derived from an estimate of the image noise (Yin and Fang, 2012; Ke et al., 2013). Finally, an even stronger assumption is that single captured and recaptured images exhibit different *JPEG (Joint Photographic Experts Group) compression artifacts* (Yin and Fang, 2012; Ni et al., 2015; Li et al., 2015). However, this assumption is only occasionally valid, for example when all images in a recapture forensics application are saved in JPEG format, so in this case single captured images are compressed once, while recaptured images can be considered as being compressed twice. When this assumption is valid, researchers derive specific features of JPEG compression artifacts to identify recaptured images (Li et al., 2015), or directly borrow features from JPEG forensics literature, *e.g.*, the one proposed in (Li et al., 2008), for detecting recaptured images (Yin and Fang, 2012; Ni et al., 2015).

2.2. Method based on image statistics

The second kind of method is based on image statistics and makes a weak assumption about the consequence of image recapture. Such methods do not aim at detecting specific and well targeted alterations that can be found in a recaptured image, but they only assume that the recapturing process introduces a subtle yet detectable deviation from the statistics of normal images. In a general sense, the only requirement is that the adopted image statistics should be

discriminative enough to reflect the statistical difference between single captured and recaptured images. For example, we can try the statistical models and features described in the monograph of Hyvärinen et al. (2009).

However, in practice, generally it is not an easy task to find a strong image statistical model that is suitable for image forensic tasks (Ng and Chang, 2013). To our knowledge, only one image-statistics-based method, from Lyu and Farid (2005), has been successfully applied for recaptured image detection. The adopted statistics features are derived in a wavelet domain. The main features are the statistical moments of wavelet coefficient prediction errors, while the moments of wavelet coefficients, without any prediction, are also included in the feature vector. This feature vector leads to relatively satisfying forensic performance for recaptured image detection, but still inferior to state-of-the-art methods which aim at detecting specific image recapture alterations, as shown in (Thongkamwitoon et al., 2015). In this manuscript, we will revisit the idea of using image statistics for recapture forensics, and we will show that such methods, if well designed and even using very simple image statistics, can reach comparable or slightly better performance than methods which attempt to detect specific image recapture alterations.

2.3. Data sets

To our knowledge, there exist two publicly available, large-scale data sets of high-resolution and high-quality recaptured images for testing and comparing methods of image recapture detection. The first data set is the database constructed by Cao and colleagues at the Nanyang Technological University (Cao and Kot, 2010; Cao, 2010) (hereafter referred to as ROSE database, with "ROSE" being the name of the authors' laboratory). It comprises 2776 recaptured images and 2710 single captured images. Among these single captured images, 2001 images were acquired by using 5 digital cameras from different makers including Canon, Casio, Lumix, Nikon and Sony. The other 709 single captured images consist of 601 high-resolution images downloaded from Internet (*e.g.*, from Flickr) and 108 high-quality tampered images. The comprised

recaptured images in ROSE database were acquired by using 3 cameras (Canon Powershot, Olympus Mju and Olympus E500) shooting displayed single captured images on 3 LCD screens (Philips 19", NEC 17" and Acer 17"). Thus there are in total 9 different camera-LCD combinations for image recapturing. The camera and environment settings (*e.g.*, camera mode, lighting and camera-to-screen distance) were manually and empirically adjusted so as to ensure a reasonably high quality of recaptured images. One interesting and important feature of ROSE database is that it includes recaptured tampered images of very high plausibility (see Fig. 1, top row, for some examples).

The second and more recent data set was constructed by Thongkamwitoon et al. (2015) from the Imperial College London (hereafter referred to as ICL database). It comprises 1035 single captured images and 2520 recaptured images, but only part of the data is freely available on-line which includes 900 single captured images and 1440 recaptured images. 9 digital cameras from 6 makers (Kodak, Nikon, Panasonic, Canon, Olympus and Sony) were used for acquiring the single captured images of diverse indoor and outdoor scenes. Recaptured images were obtained by using 8 cameras from 5 makers (Panasonic, Nikon, Canon, Olympus and Sony) which shot displayed single captured images on an NEC 23" LCD screen. Among these 8 cameras used for image recapture, 5 of them were previously used to acquire single captured images included in the same data set. In general, the recaptured images in the ICL database are of even higher quality when compared to those in the ROSE database. The reason is that the recapturing parameters (in particular camera-to-screen distance and camera's lens aperture) were carefully determined with convincing theoretical justification, so that aliasing distortions have been largely reduced to a nearly invisible level (Thongkamwitoon et al., 2015) (see the original paper for details). This is also the most striking feature of the ICL database.

We can see that both data sets include a large number of high-resolution and high-quality single captured and recaptured images for which a variety of digital cameras from different makers were used as image acquiring devices. The authors of the ROSE database also used multiple LCD screens from different man-

ufacturers for image recapturing. The main difference between ICL and ROSE databases resides in the adopted recapturing technique, which results in quite different level of aliasing artifact in the recaptured images. Therefore, we can consider that images comprised in either data set are comprehensive enough to represent images acquired by various digital cameras, either in a single captured or recaptured setting. Furthermore and intuitively, single captured images from the two data sets should have more or less similar statistical properties; however, it can be expected that there is noticeable statistical difference between recaptured images from the ICL and ROSE databases since the adopted recapturing technique differs, which leads to different properties of the induced distortion. Later in this manuscript, we will test our proposed method on these two popular databases and conduct comparisons with existing methods including the two very recent algorithms from Li et al. (2015) and Thongkamwitoon et al. (2015). The corresponding results, obtained under intra-database, inter-database and combined-database settings, demonstrate the effectiveness of our method and provide us with useful insights for future working directions. More precisely, we show that the assumption and the basic idea of our method (*i.e.*, image recapturing, though possibly with slight technical differences, will introduce subtle yet detectable image statistical alterations) are valid on the individual ICL and ROSE database as well as on a combined version of the two data sets.

3. Proposed Method

In this section, we will begin with a brief presentation of the motivations and an overview of the proposed method for detecting recaptured images in Section 3.1. Then, in Section 3.2 we provide the technical details of our method and attempt to explain the intuitions behind the design of the adopted image statistics feature.

3.1. Motivations and overview

We have observed two current trends in the research of recaptured image detection. The first trend is to combine features for detecting various alterations

that may be introduced by image recapture, as in (Ke et al., 2013; Ni et al., 2015; Li et al., 2015). The combined feature is expected to be able to detect all the considered alterations and thus has more chance to accurately classify between single captured and recaptured images. In fact, combining features is not a new idea, in the first influential work on recaptured image detection of Cao and Kot (2010), the authors already combine texture, blurriness and color descriptors together to form a more powerful feature. The main drawbacks of such methods are the potentially high computational complexity and the high dimensionality of the combined feature.

The second trend is to make use of advanced machine learning tools to solve this classification problem between single captured and recaptured images, in the hope of further boosting the forensic performance. The recent work of Thongkamwitoon et al. (2015) uses sparse dictionary learning to obtain a distinctive feature of edges in the two kinds of images. This can actually be considered as an effective representation learning procedure to automatically obtain appropriate and distinctive features from single captured and recaptured images. This dictionary-learning-based method has very good performance for detecting recaptured images in which blurriness is the dominant alteration. However, the sparse dictionary training is very time consuming according to a personal communication with the first author of (Thongkamwitoon et al., 2015), even on a limited number of training samples. In addition, there are quite a few parameters whose values are empirically fixed in particular for the edge extraction algorithm, and this, along with other factors, could lead to performance drop on certain cases. Indeed, as demonstrated through our experiments presented in Section 4, the edge profile dictionaries trained on full-sized images appear quite sensitive and give poor performance on cropped images from full-sized ones.

It is worth mentioning that all the methods following the two trends aim at detecting one or several specific alterations due to image recapture. By contrast, image-statistics-based methods do not make use of such strong prior knowledge and are based on, to some extent, a rather generic statistics feature. Therefore, this kind of methods is expected to be conceptually simple and computationally

efficient, without any heterogeneous feature combination or complicated learning procedure. In addition, image-statistics-based methods have more chance to be capable of producing equally good results on either full-sized or cropped images, if the adopted image statistics feature is well designed and strong enough to describe the intrinsic difference between single captured and recaptured images.

For image-statistics-based methods, it is somehow surprising to see that researchers focus on relatively complex statistical features in a transformed wavelet domain (Lyu and Farid, 2005). In this manuscript, we would like to go back to the origin of the idea of using image statistics for recapture forensics and attempt to accomplish this forensic task by using very simple yet well motivated image statistics derived directly in the pixel domain. More precisely, for a given image, we extract a feature vector consisting of simple pixel-wise correlation coefficients in image differential domains. The extracted feature vector will be demonstrated later to convey discriminative information for the task of recaptured image detection. Figure 2 illustrates the block diagrams of the training and testing stages of the proposed forensic method. During the off-line training, first of all, the aforementioned image statistics feature is extracted from a group of single captured and recaptured images. Then, the extracted feature vectors, along with the associated image labels (*e.g.*, "0" for single captured ones and "1" for recaptured ones), are fed into a Support Vector Machine (SVM) (Cortes and Vapnik, 1995) for training a recapture forensic detector. During the on-line testing, for a given image, a feature vector is extracted from it and input to the trained SVM classifier in order to get the forensic result of whether the given image is a single captured or a recaptured one. It is worth noting that in our method, with a simple and quick feature extraction algorithm, forensic testing on one image is computationally very efficient and can be done within a few seconds. In the meanwhile, we can ensure that the off-line training stage can be accomplished within a reasonable amount of time. These will be demonstrated in Section 4 through comparative experimental results.

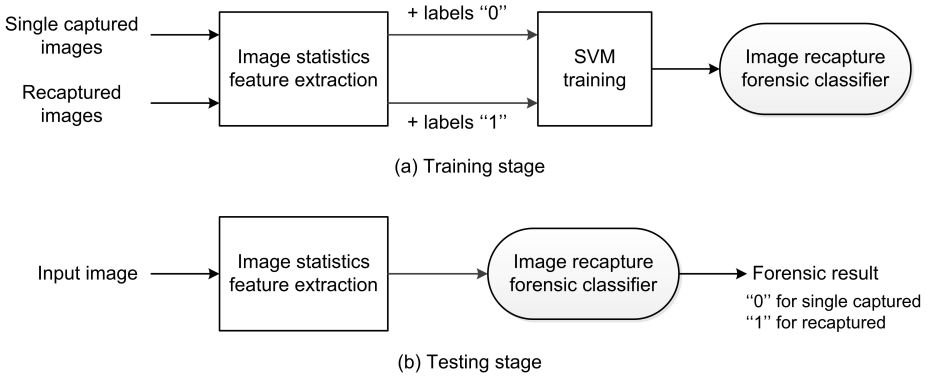

Figure 2: Block diagram of (a) the off-line training stage and (b) the on-line testing stage of the proposed image recapture forensic method.

3.2. Image statistics feature

From the last subsection on the motivations of our work, we can summarize the guide lines for the design and implementation of the to-be-proposed image statistics feature as follows: The adopted feature should be discriminative enough to reflect the difference between the two kinds of images under consideration, but should also be simple enough to be extracted from a given image. In order to satisfy both constraints, we choose to compute simple statistics in differential pixel domains, as detailed below.

3.2.1. Working in image differential domains

Concerning the requirement of being discriminative, ideally the feature should not be dependent on the content of the individual image, but only sensitive to the difference between single captured and recaptured images. To this end, a common strategy is to remove the image's low-frequency component and in consequence to make assumption that the discriminative information is hidden in the high-frequency part of the image. This was actually the strategy followed by Lyu and Farid (2005) who extracted image statistics features from high-frequency wavelet subbands. Different from their method which works in a transformed domain, we choose to work directly in the pixel domain. This is in part motivated by practical observations from the image forensics research, that

is, when we want to detect an image modification introduced in a certain domain (*e.g.*, in the pixel domain or in a transformed domain), it would be safe and often advantageous to work in the same domain as the modification operation. For instance, when a forensic analyst intends to detect contrast enhancement via spatial-domain Gamma correction, it is very effective to work directly in the pixel domain by studying the pixel value histogram (Stamm and Liu, 2010); by contrast, when we want to expose DCT (Discrete Cosine Transform) coefficients quantization artifacts and even estimate the quantization step, it would be better to work in the DCT domain by investigating the DCT coefficients statistics (Fan and de Queiroz, 2003). In our case, we consider that recapturing digital images from LCD screens is an operation that introduces alterations essentially in the spatial domain; at least it does not explicitly work in a transformed domain. Therefore, we have the intuition that it would be beneficial to work directly in the pixel domain when attempting to derive a discriminative image statistics feature for recapture forensics.

More precisely, for a given image $\mathbf{X}$ of size $M \times N$, we first of all apply low-pass filtering on it using two simple filters, then we compute two residue images, *i.e.*, the difference between $\mathbf{X}$ and its low-pass filtered version. Formally, the two residue images $\mathbf{R}^{(i)}, i \in \{1, 2\}$ are calculated as follows:

$$\mathbf{R}^{(i)} = trim\left(\mathbf{X} - \mathbf{X} * \mathbf{f}^{(i)}\right), \tag{1}$$

where $*$ means mathematical convolution and the function $trim(.)$ removes the first row and column, as well as the last row and column, from the input image (explained later in the next paragraph). The two convolution kernels $\mathbf{f}^{(1)}$ and $\mathbf{f}^{(2)}$ are given as

$$\mathbf{f}^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{f}^{(2)} = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{bmatrix}. \tag{2}$$

We can see that for a given pixel x_j in $\mathbf{X}$, the corresponding pixel value in the residue image $\mathbf{R}^{(1)}$ (respectively $\mathbf{R}^{(2)}$) measures the difference between the

value of x_j and the average value of its two horizontal (respectively vertical) neighboring pixels in $\mathbf{X}$. These operations, to some extent, remove the image's low-frequency component that is much dependent on the image's content but not discriminative for the task of recapture forensics. From another point of view, the residue images basically describe the edges and noises in the image, respectively more or less related to the blurriness and aliasing-like alterations, and they might have quite different characteristics for single captured and recaptured images. Therefore, from the residue images it would be possible to extract a statistical feature which can expose multiple alterations in a recaptured image. One detail related to the boundary condition of the filtering is that we only consider pixels in $\mathbf{X}$ that have complete neighboring pixels as required by the filters given in Eq. (2). Therefore, the final residue images will have two rows and two columns less than the original image, as reflected by the $trim(.)$ function in Eq. (1). We have attempted to use other filters (*e.g.*, filters measuring the mean of central pixel's 4-connected von Neumann neighbors and its 8-connected Moore neighbors) and have found that the two simple filters $\mathbf{f}^{(1)}$ and $\mathbf{f}^{(2)}$ yield slightly better forensic performance than others. Therefore, we decided to use the two filters as given in Eq. (2) in the proposed image recapture forensic method. A related discussion on filters, which may open interesting future working directions, can be found later in Section 4.4.

The bottom row of Fig. 3 illustrates the visual appearance of the residue images (after proper post-processing as explained in the figure's caption) for a pair of single captured and recaptured images from ICL database. We can see that the two residue images exhibit very different characteristics. First, in general, in textured regions, the residue of a recaptured image has lower-amplitude pixels than that of a single captured image, mainly due to the blurriness alteration (see and compare the regions corresponding to the trees and the temple). Second, although in ICL database the aliasing-like distortion is well controlled at an invisible level, such distortion can still be exposed in the residue of a recaptured image (notice the periodic, strip-shaped pattern in the sky part). In all, Fig. 3 implies that for the purpose of image recapture forensics, it would

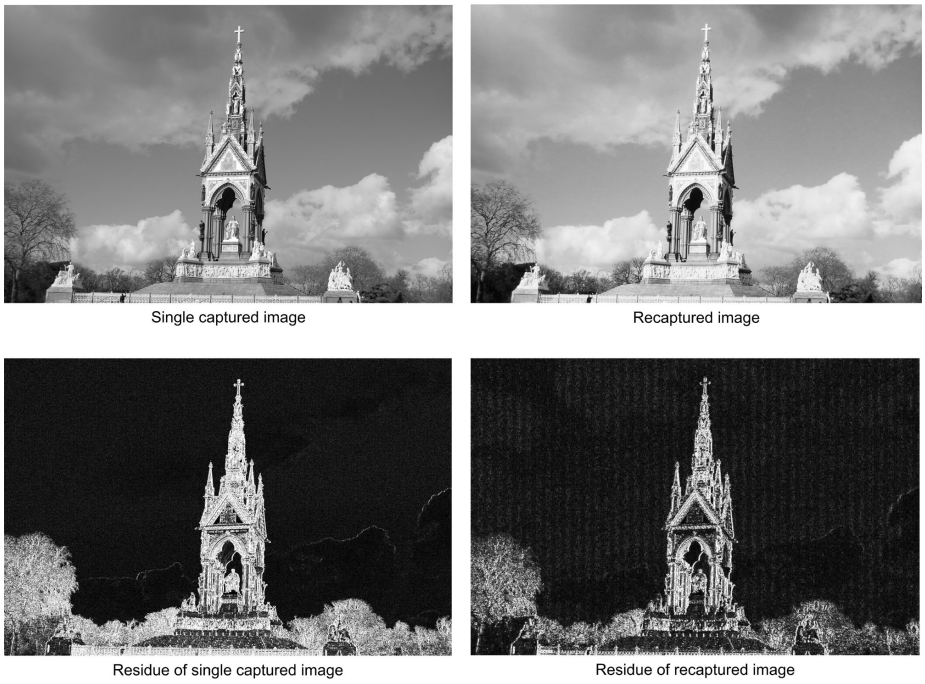

Figure 3: In the top row we show a pair of single captured and recaptured images from ICL database. Following Thongkamwitoon et al. (2015), we have converted the original color images to grayscale before forensic analysis. In the bottom row are the corresponding residue images $\mathbf{R}^{(1)}$ of the two images in the top row. The two residue images have very different characteristics due to alterations introduced by image recapture. For a better visualization, we have taken pixel-wise absolute values of the residue images and performed thresholding at 15 (the range of pixel values in unfiltered images is from 0 to 255).

be relevant to derive an effective image statistics feature in pixel differential domains as given by the residue images.

3.2.2. Pixel-wise correlation coefficient

The statistics feature is computed from $\mathbf{R}^{(i)}, i \in \{1, 2\}$ and the computation is exactly the same for the two residue images. We keep in mind the requirements of simplicity and discriminative capability of the feature and the fact that we often need to find a good trade-off between the two requirements. The simplest statistics of $\mathbf{R}^{(i)}$ is the marginal distribution of residue values at individual pixels. However, as pointed out in (Ng and Chang, 2013), in general

such simple marginal statistics is a weak feature and cannot provide satisfying forensic performance. Therefore, we need to use and compute a statistics feature somehow related to the joint distribution of values at different pixels in a residue image. In this case, it is a natural idea to start with the pixel-wise *correlation coefficient* (CC) between the residues.

More precisely, from a residue image $\mathbf{R}^{(i)}$, we first of all extract all its overlapping 5×5 patches denoted by $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, ..., \mathbf{P}^{(K)}$ with K the total number of extracted patches.[1] The elements in each patch $\mathbf{P}^{(k)}$ are denoted by $p_{i,j}^{(k)}$ with $i, j \in \{1, 2, 3, 4, 5\}$ and $k \in \{1, 2, ..., K\}$. Then, we pick up the residue value with the same index i, j from each patch and concatenate them together to form a vector as $\mathbf{V}_{i,j} = \left[p_{i,j}^{(1)}, p_{i,j}^{(2)}, ..., p_{i,j}^{(K)}\right]$, and in all we have 25 such vectors. Later, we compute the correlation coefficient $c_{i,j}$ between the vector corresponding to the central pixel, *i.e.*, $\mathbf{V}_{3,3}$, and all the 25 vectors. Specifically, we have

$$c_{i,j} = \rho_{\mathbf{V}_{3,3}, \mathbf{V}_{i,j}} = \frac{\sum_{k=1}^{K} \left(p_{3,3}^{(k)} - \bar{p_{3,3}}\right) \left(p_{i,j}^{(k)} - \bar{p_{i,j}}\right)}{\sqrt{\sum_{k=1}^{K} \left(p_{3,3}^{(k)} - \bar{p_{3,3}}\right)^2} \sqrt{\sum_{k=1}^{K} \left(p_{i,j}^{(k)} - \bar{p_{i,j}}\right)^2}}, \tag{3}$$

where $\bar{p_{3,3}}$ and $\bar{p_{i,j}}$ are respectively the mean of the elements in $\mathbf{V}_{3,3}$ and $\mathbf{V}_{i,j}$, and we have $i, j \in \{1, 2, 3, 4, 5\}$.

In practice, for each residue image $\mathbf{R}^{(1)}$ and $\mathbf{R}^{(2)}$, we can compute 25 correlation coefficient values and regroup them in a 5×5 matrix with i and j as respectively the row and column index. The CC values for the two residue images are denoted by $c_{i,j}^{\rightarrow}$ and $c_{i,j}^{\downarrow}$ indicating that they correspond to respectively horizontal and vertical residues. These CC values measure how the horizontal or vertical residues, at different locations within a small local neighborhood of size 5×5, are related to each other through second-order mixed statistical moments. We expect that the extracted CC values are sensitive to the alterations

[1] For the sake of simplicity, we drop, in the notation of patches, the index of residue image. The feature extraction procedure is in fact the same for the two residue images. A brief discussion on the influence of the patch size can be found later in Section 3.2.3 and some relevant experimental results can be found in Section 4.1.

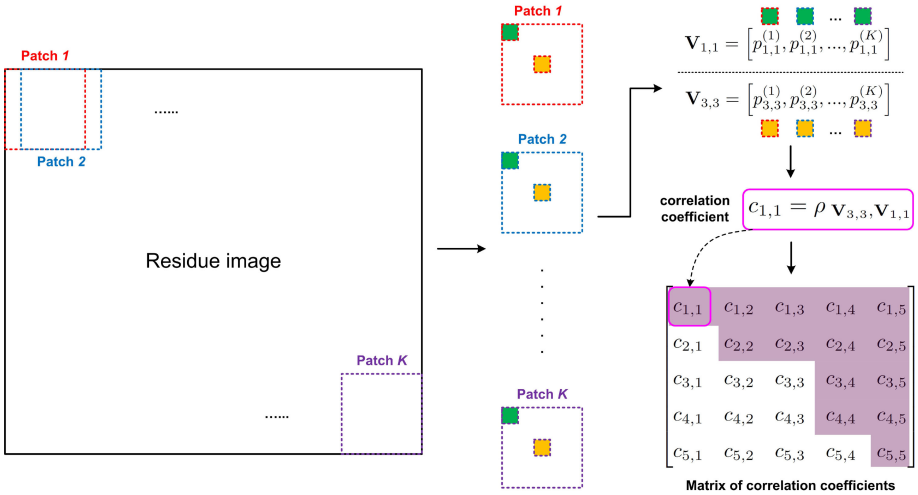

Figure 4: Illustration of the patch extraction and the computation of pixel-wise correlation coefficients, taking $c_{1,1}$ as example. The shaded elements in the matrix of correlation coefficients are retained elements comprised in the proposed feature vector.

due to image recapture and exhibit different characteristics for single captured and recaptured images. Figure 4 illustrates the procedure of extracting a matrix of correlation coefficients from a residue image.

We show, for ICL database, the average CC value matrix of $c_{i,j}^{\rightarrow}$ of all the released 900 single captured images in Eq. (4) and that of 1440 recaptured images in Eq. (5).

$$\bar{\mathbf{M}}_{single}^{(ICL)} = \begin{bmatrix} -0.1697 & 0.0066 & 0.3743 & 0.0021 & -0.1688 \\ -0.3104 & -0.0327 & 0.6965 & -0.0338 & -0.3112 \\ -0.4237 & -0.0972 & 1.0000 & -0.0972 & -0.4237 \\ -0.3112 & -0.0338 & 0.6965 & -0.0327 & -0.3105 \\ -0.1688 & 0.0021 & 0.3743 & 0.0066 & -0.1697 \end{bmatrix}. \tag{4}$$

$$\bar{\mathbf{M}}_{recapture}^{(ICL)} = \begin{bmatrix} -0.2459 & 0.2079 & 0.5853 & 0.2160 & -0.2358 \\ -0.3348 & 0.2435 & 0.8129 & 0.2515 & -0.3299 \\ -0.3834 & 0.2253 & 1.0000 & 0.2254 & -0.3834 \\ -0.3299 & 0.2515 & 0.8129 & 0.2435 & -0.3348 \\ -0.2358 & 0.2160 & 0.5853 & 0.2079 & -0.2460 \end{bmatrix}. \tag{5}$$

The difference between the mean CC values extracted from the two kinds of images implies the strong discriminative capability of the proposed image statistics feature (see and compare the elements with the same index in the two matrices). More precisely, the CC values between the central pixel and most neighbors in the local 5×5 patch are noticeably higher for recaptured images than those for single captured images, in particular for the elements in the second, third and fourth columns of the two matrices. This is understandable because $c^{\rightarrow}_{i,j}$ are computed from the horizontal residue image $\mathbf{R}^{(1)}$, and actually we have the same observation for elements in the second, third and fourth rows in the matrices of $c^{\downarrow}_{i,j}$ computed from the vertical residue image $\mathbf{R}^{(2)}$ (for the sake of brevity, here we do not show the results of $c^{\downarrow}_{i,j}$). This CC value increase in recaptured images is probably due to the introduced blurriness effect. Another interesting observation is that the CC values between the central pixel and distant neighbors, in particular those at the four corners of the CC value matrix, in recaptured images are lower than those in single captured images, see and compare the corner elements in Eqs. (4) and (5). This CC value decrease is probably caused by the aliasing-like distortion that is still present, although invisible, in the recaptured images (see Fig. 3, right column). We think that in this case the high-frequency aliasing-like distortion can reduce the similarity between certain pairs of residues at distant pixels.

As shown above from the results on the ICL database in Eqs. (4) and (5), we observe a strong symmetry in the extracted matrix of CC values. In order to remove redundancy and to reduce the negative impact of redundant features during the SVM training, we only retain the CC values in the upper triangle of the matrix. In addition, we also remove the central element from the matrix which is always equal to 1 thus non-discriminative for the forensic classification. The retained correlation coefficients are illustrated as shaded elements in the matrix of correlation coefficients on the bottom right corner of Fig. 4. In consequence, we can obtain a 14-dimensional feature vector from each of the two residue images. Therefore, the dimension of the final, concatenated feature vector is $14 \times 2 = 28$.

As mentioned in Section 3.1 and shown in Fig. 2, this 28-dimensional feature is extracted from a large number of images (with ground-truth labels) for the purpose of training the SVM-based forensic classifier, and later from one or several test images (with unknown labels for the trained SVM-based classifier) before forensic testing. The low dimensionality of the feature vector, along with its computational simplicity, allows us to efficiently extract features and perform large-scale training and testing.

3.2.3. Discussion on technical choices

We choose to compute the correlation coefficient instead of the straightforward correlation still for the concern of being independent of the image's content. In fact, it is easy to see that the straightforward correlation value depends on the overall brightness of the image: Two images of the same scene but of different brightness lead to quite different correlation values between residues. By contrast, the computation of correlation coefficient incorporates a kind of divisive normalization procedure, which makes the CC value independent of the amplitude of the input vectors.

We have also tried to build forensic detectors using patches of size 3×3 and 7×7 pixels. It is found that 3×3 patches lead to a slight performance drop compared to 5×5 patches, probably due to the decrease of discriminative capacity caused by a smaller local neighborhood. In the meanwhile, 7×7 patches result in a feature vector of almost twice of the dimensionality of that of 5×5 patches, but the forensic performance remains more or less comparable. One exception is the recapture forensics of very high-resolution images (*e.g.*, with a width of 4096 pixels), for which 7×7 neighborhood has about 0.5% higher forensic accuracy than 5×5 neighborhood. This however can be understood, because in a very high-resolution image, neighboring pixels tend to have very high similarity and it would be better to consider pixels in a larger local window for better illustrating the statistical difference between single captured and recaptured images. Nevertheless, in practice the 7×7 neighborhood has a considerable overhead of computational cost for extracting features from 7×7

patches (especially on very high-resolution images) and for SVM training due to dimensionality increase. Considering all the above points, we choose 5×5 as the size of overlapping patches for feature extraction. Concrete experimental results regarding different patch sizes will be given in Section 4.1.

More sophisticated statistical features have been considered during algorithm design and implementation, including higher-order mixed statistical moments such as coskewness and cokurtosis, as well as the multiple correlation coefficient between three or more variables (Abdi, 2007). However, we have observed that the computation of higher-order statistics is much more time-consuming than computing correlation coefficient while their inclusion in the feature vector does not lead to noticeable performance improvement. We have the same observation for the multiple correlation coefficient. Hence, in order to keep a good balance between algorithm simplicity and forensic performance, we only use correlation coefficients as the image statistics feature in our method.

4. Experimental Results and Discussion

In this section, we will present the experimental results of the proposed method on two large databases of high-resolution and high-quality recaptured images from LCD screens. We will also provide comparison results between our method and a number of existing methods, including the very recent methods of Thongkamwitoon et al. (2015) and Li et al. (2015). At the end of this section, we will present some discussions on the proposed method, which might inspire new ideas of more effective methods for recaptured image detection. The proposed method has been implemented in Matlab®, and the source code of the statistics feature extraction is freely shared on-line[2].

4.1. Experiments on ICL database

We will first of all focus on the experimental validation on ICL database which, compared to ROSE database, is a more recent and more challenging

[2]Available at `http://www.gipsa-lab.fr/~kai.wang/recap_feature_extraction.m`

database and comprises higher-quality recaptured images with almost invisible aliasing distortions. The experiments were conducted on the released 900 single captured and 1440 recaptured images. We will mainly compare our method with the method of Thongkamwitoon et al. (2015), to our knowledge the most effective recapture forensic method on this database. In order to ensure a fair comparison with their method, we have followed as closely as possible the experimental setting described in their original paper. More precisely, following (Thongkamwitoon et al., 2015), we first converted the color images comprised in the database to grayscale[3], and then all the grayscale images were rescaled while keeping the ratio between height and width unchanged so that the resized version has a width of 2048 pixels. Still following (Thongkamwitoon et al., 2015), we keep the ratio between the number of images in the training set and that in the testing set as 15:100. In the SVM-based classifier, we chose to use the popular and effective RBF (Radial Basis Function) kernel. In addition, following common guide lines from applied machine learning, the values of SVM hyper-parameters were determined by using 5-fold cross validation on the training set. Once the SVM-based forensic classifier is trained by using labeled images in the training set, we test and report its forensic classification performance on the "unseen" images in the testing set. The forensic performance is evaluated by using the classification accuracy (*i.e.*, the percentage of correctly classified images) of single captured images, recaptured images and all the images in the testing set. In order to enhance the statistical significance of the obtained results, for each test of the proposed method we performed 50 runs, and for each run we had a different and randomly partitioned training and testing sets while always keeping the ratio between the number of training and testing images unchanged. At the end, we report the classification accuracy as the mean of the

[3]We use the Matlab® `rgb2gray` function for converting color images to grayscale. Other methods, possibly with different technical details (*e.g.*, using different weighting factors for three color channels), might be used. However, in order to reproduce results as close as possible to those shown in this manuscript, we recommend using the `rgb2gray` function.

results obtained from the 50 runs.

In the following, we will first of all show experimental results relative to different patch sizes considered for the computation of pixel-wise correlation coefficients. We have tested the forensic accuracies for 3×3, 5×5 and 7×7 patches and the obtained results are presented in Table 1, under different image resolutions (values in the first column) and along with the dimensionality of the corresponding feature vector for patches of different sizes (given in parentheses of the second column). Here the image resolution means the width of rescaled images before feature extraction mentioned in the last paragraph, and the considered values of this width include 2048 (value used by Thongkamwitoon et al. (2015)), 3072 and 4096 pixels. As mentioned above, the classification accuracy reported in Table 1 is the mean of the results obtained from 50 runs with random partitioning of training and testing sets. The highest forensic accuracies under different image widths are highlighted in bold. From this table, it can be seen that 5×5 and 7×7 patches always give very satisfying forensic results with overall classification accuracy all higher than 97.5% and sometimes as high as 99%. A general observation is that when the width increases to 3072 and 4096 pixels, the performance decreases for 3×3 patches but increases for 5×5 and 7×7 patches. This is somehow expected because as the width increases neighboring pixels tend to have an enhanced similarity; therefore in this case, a relatively large patch size would be beneficial to comprise discriminative statistics features for distinguishing between single captured and recaptured images.

In all, it appears that 3×3 patches give acceptable results but are not big enough to incorporate very strong and discriminative features for recapture forensics, while 5×5 and 7×7 patches both yield very good results under all the considered width values. When we make a choice between 5×5 and 7×7 patches, we also consider their computational cost. As mentioned earlier, 7×7 patches result in higher-dimensional features as well as more costly feature extraction and SVM training. For example, the feature extraction on 7×7 patches is about two times slower than that on 5×5 patches. Therefore, in order to have a good trade-off between forensic accuracy and computational cost, we choose to use

Table 1: Comparison of feature dimensionality and classification accuracies (on single captured images, recaptured images and all the testing images in ICL database) of our proposed forensic detector under different image widths and using different patch sizes. Below, "dim." stands for "dimensionality" and "accu." stands for "accuracy". The highest accuracy (single, recapture or overall) under a fixed image width is highlighted in bold.

Image width	Patch size (Feature dim.)	Accu. single	Accu. recapture	Overall accu.
2048 pixels	3×3 (10)	**96.96**%	97.80%	97.47%
	5×5 (28)	96.27%	**98.62**%	**97.71**%
	7×7 (54)	96.23%	98.38%	97.55%
3072 pixels	3×3 (10)	96.46%	96.94%	96.76%
	5×5 (28)	98.59%	99.08%	98.89%
	7×7 (54)	**98.72**%	**99.33**%	**99.09**%
4096 pixels	3×3 (10)	93.60%	98.11%	96.37%
	5×5 (28)	98.28%	98.77%	98.58%
	7×7 (54)	**98.99**%	**99.01**%	**99.01**%

5×5 patches in all subsequent experiments. It is worth pointing out that for the sake of a fair comparison with the method of Thongkamwitoon et al. (2015), we will use the image width of 2048 pixels on the ICL database (under this width 5×5 patches actually give the best overall accuracy of 97.71%), although when we use a bigger width, our method, with 5×5 and 7×7 patches, has even higher overall classification accuracies (all higher than 98.5%). It can also be noticed that the overall classification accuracies of all the considered patch sizes and under the width of 2048 pixels are higher than that of the state-of-the-art method of Thongkamwitoon et al. (2015) (compare the results given in Tables 1 and 2). This implies the high discriminative capability of the proposed image statistics feature.

Then we compare our method with a number of representative existing methods, including the pioneer method from Lyu and Farid (2005) that is based on images statistics features extracted from the wavelet domain, the method from Cao and Kot (2010) which combines multiple features, and the one from

Thongkamwitoon et al. (2015) which attains until now the highest forensic accuracy on ICL database. The comparison results are presented in Table 2, from which we can see that our method gives the best overall classification accuracy, slightly higher than that of the state-of-the-art method of Thongkamwitoon et al. (2015). Our method significantly outperforms the method of Lyu and Farid (2005), which implies that for image recapture forensics it might be more suitable to extract image statistics features directly from the spatial domain rather than from the wavelet domain (also see the discussion in Section 3.2.1). The feature-combination-based method of Cao and Kot (2010) does not perform very well on ICL database. As analyzed in (Thongkamwitoon et al., 2015), this may be due to the fact that one of the features in the method of Cao and Kot (2010) is designed specifically for detecting aliasing alterations, while recaptured images in ICL database do not have very obvious aliasing-like distortions. Therefore, in this case, this specific feature might become not that discriminative and result in a decrease in classification accuracy. At last, when compared with the method of Thongkamwitoon et al. (2015), our method has a slightly lower accuracy on recaptured images but a higher accuracy on single captured images, leading to a slightly higher overall accuracy than their method (97.71% vs. 97.44%). This is a quite positive result considering the following two facts: First, our image-statistics-based method makes a weak assumption on the image recapture process while the method of Thongkamwitoon et al. (2015) makes use of strong prior knowledge of the blurriness effect present in the recaptured images; Second, our method has other good properties and advantages as detailed in the following paragraphs.

Besides the slightly higher overall classification accuracy, when compared to the state-of-the-art method of Thongkamwitoon et al. (2015), our method is also computationally more efficient and has better forensic performance on cropped images from full-sized ones. We will in the first place present the experimental results concerning the computational efficiency. In practical forensic scenarios, it is important for a forensic detector to be fast in both the on-line testing stage and the off-line training stage. The on-line testing stage is typically composed of two

Table 2: Classification accuracies of different forensic methods on single captured images, recaptured images and all the test images in ICL database. Below, "accu." stands for "accuracy". The highest accuracy in each column is highlighted in bold. The results in the second to fourth rows are extracted from (Thongkamwitoon et al., 2015).

Method	Accu. single	Accu. recapture	Overall accu.
Lyu and Farid	87.56%	90.04%	89.09%
Cao and Kot	83.67%	92.02%	88.81%
Thongkamwitoon et al.	94.89%	**99.03**%	97.44%
Ours	**96.27**%	98.62%	**97.71**%

steps, *i.e.*, feature extraction and the forensic classification. Table 3 shows, for our method and the method of Thongkamwitoon et al. (2015), the comparison results of the execution time of the on-line testing stage. The presented values in this table are means of results collected from 200 images including 100 single captured ones and 100 recaptured ones, and the experiments were conducted on a laptop equipped with an Intel® i5 2.27 GHz CPU and 4 GB RAM. The execution time of the method of Thongkamwitoon et al. (2015) were obtained by using the authors' Matlab® source code shared on the Internet. From Table 3, we can see that our method is significantly faster, with an execution time per image of about 2.93 seconds versus about 12.74 seconds for their method. It is difficult to quantitatively compare the execution time of the training stage as the dictionary learning code of (Thongkamwitoon et al., 2015) is not available, but it is safe to say that qualitatively our method is much faster than their method. More precisely, there are 304 images used for training and for our method the feature extraction from all these images takes about 15 minutes and the SVM training spends about 4 to 5 minutes, which leads to a total training time of about 20 minutes. By contrast, for the method of Thongkamwitoon et al. (2015), the feature extraction from 304 training images already takes about 50 minutes, much longer than the total training time of our method. If we further include the time-consuming step of dictionary learning, the training time of their method would be even longer. In all, it can be seen that our method is computationally

Table 3: Comparison of the average execution time per image (in seconds) of the on-line testing stage between our method and the method of Thongkamwitoon et al. (2015).

Method	Feature extraction	Classification	Total
Thongkamwitoon et al.	10.2218	2.5226	12.7444
Ours	2.9303	0.0013	2.9316

much more efficient, and this efficiency is mainly due to the conceptual and algorithmic simplicity of the proposed feature and its extraction.

Another advantage of our method is its good performance when trained on full-sized images and later tested on cropped images of full-sized ones. Cropping is in fact commonly involved in the generation of recaptured tampered images. For instance, the cropping operation is necessary to remove the LCD frame (if any) from a recaptured image. It is also possible that creators of tampered images crop a specific tampered part from a full-sized image, either to eliminate visual clues present in the removed part that can be telltale of the image falsification, or to make the final tampered image more focused and more plausible. Under this context, it would be beneficial if a forensic detector, after being trained on full-sized images, can provide reliable forensic results on cropped testing images. This usually requires a good stability of the feature under full-sized and cropped images. Our method actually has such a good property. Table 4 presents, for our method and the method of Thongkamwitoon et al. (2015), the testing results on cropped images extracted from the center of the full-sized testing images in ICL database. We can see that the performance of our method decreases gradually and gracefully as the cropping size decreases and remains very satisfying for cropped images of size down to 256×256 pixels, which demonstrates the stability and discriminability of our feature. By contrast, the method of Thongkamwitoon et al. (2015) is very sensitive to cropping, probably due to the limited generalization capability of the trained dictionaries and the specific parameter setting used for edge extraction.

Table 4: Comparison of the forensic accuracy on cropped images of different sizes between our method and the method of Thongkamwitoon et al. (2015).

Method	Size of cropped images			
	1024×1024	512×512	256×256	128×128
Thongkamwitoon et al.	61.52%	61.89%	62.74%	54.10%
Ours	97.37%	95.97%	93.22%	87.75%

4.2. Experiments on ROSE database

In this subsection, we briefly present the experimental results obtained on ROSE database, a larger but seemingly less challenging database than ICL database. ROSE database comprises 2710 single captured images and 2776 recaptured images. All the included images are of high resolution and high visual quality, except for the fact that on one part of the recaptured images we can observe slightly visible aliasing-like distortions. In general, the aliasing artifacts in ROSE database is stronger than those in ICL database, which makes ROSE database somehow less challenging for forensic detection. However, one striking characteristic of ROSE database is that it comprises a large number of recaptured tampered images (some examples can be found in the top row of Fig. 1). In order to convincingly illustrate that image recapture is indeed helpful in enhancing the plausibility of a tampered image (also refer to explanations given in Section 1), we show in Fig. 5 close-ups of an authentic image, a falsified image, and the corresponding recaptured falsified image from an LCD screen. It can be seen that image recapture is indeed useful to hide the unnatural and sharp transition near the border of a spliced subimage.

It is interesting to evaluate the performance of our method on ROSE database, in order to see whether it can cope well with different kinds of alterations in recaptured images (*e.g.*, blurriness and weak or strong aliasing) and to check its potential in practical applications of detecting recaptured falsified images. We will compare our method with that of Li et al. (2015), the most recent method that reports detailed and very good results on ROSE database. Their method is however based on strong assumptions of increased JPEG compression

(a) (b) (c)

Figure 5: Close-ups of (a) an authentic image, (b) a tampered image in which a street lamp has been spliced into the authentic image, and (c) the corresponding recaptured tampered image from an LCD screen. The full-sized image of (c) is shown in Fig. 1, top-right corner. We can more or less perceive unnatural transition near the border of the street lamp in (b) while the close-up image in (c) is visually much more plausible. The blurriness effect in (c) is in fact quite similar to "natural" blurriness, *e.g.*, that introduced by out of focus of the camera or hand trembling of the photographer when taking the picture.

artifacts and existence of aliasing-like distortions in recaptured images. When conducting the experiments, we followed the description in (Li et al., 2015) and kept the ratio between the number of training and testing images as 5:1. We have converted the color images in the database to grayscale and resized the grayscale images to have a fixed width of 3072 pixels (instead of 2048 pixels for ICL database because a large number of images in ROSE database have a width very close to 3072 pixels and this technical choice leads to a slightly higher overall accuracy). The SVM hyper-parameters were determined by using 10-fold cross validation on training set, and we report mean forensic accuracies of 10 runs on testing data, with different and randomly portioned training and testing data sets for each run. Compared to ICL database, we increase the number of folds and decrease the number of runs mainly because ROSE database comprises more images than ICL database. Table 5 presents the experimental results on ROSE database. In all, our method gives very good classification accuracy results that are slightly higher than those reported in (Li et al., 2015). This illustrates the very good discriminability of the proposed statistics feature on different databases and with regard to different situations of image recap-

Table 5: Classification accuracies of the method of Li et al. (2015) and our method on ROSE database. Below, "accu." stands for "accuracy". The results in the second row are extracted from (Li et al., 2015).

Method	Accu. single	Accu. recapture	Overall accu.
Li et al.	98.63%	99.27%	98.95%
Ours	99.00%	99.35%	99.18%

ture alterations, as well as the practical applicability of our method in detecting recaptured tampered images.

4.3. Experiments on both databases

In this section, we will present the experimental results of our method on images from both databases, either in an inter-database setting or in a combined-database setting. We first consider the very challenging scenario of inter-database setting, meaning that we train a recapture forensic detector on one database and test the performance of the trained detector on the other database. In this series of experiments, in order to keep consistency, prior to feature extraction all the images in ICL and ROSE databases are rescaled to have a width of 2048 pixels. The second and third rows of Table 6 present the results of the two specific setups of this series of inter-database experiments, *i.e.*, we train a classifier on the ICL database and test it on the ROSE database (second row) and vice versa (third row). It can be seen that the performance is in general poor, with a low overall classification accuracy. The obtained results are indeed interesting, with a relatively high accuracy in detecting single captured images and a very low accuracy for detecting recaptured images. In particular, when a detector is trained on ROSE database and tested on ICL database, we can obtain a nearly perfect classification on single captured images and an almost 0% classification accuracy on recaptured images. Intuitively, the former can be attributed to the fact that the ROSE database comprises a large amount of diverse single captured images, and these images are involved in the classifier training and will help the classifier accurately identify this kind of images even

Table 6: Classification accuracies of our method in inter-database and combined-database settings. Below, "accu." stands for "accuracy". The pattern "databaseA → databaseB" means that the classifier is trained on "databaseA" and later tested on "databaseB". Therefore, the second and third rows present results in the inter-database setting, while the last row presents results in the combined-database setting.

Experimental setting	Accu. single	Accu. recapture	Overall accu.
ICL → ROSE	80.34%	46.42%	63.18%
ROSE → ICL	99.14%	0.22%	38.26%
ICL+ROSE → ICL+ROSE	96.54%	98.16%	97.41%

on a different data set. The latter can be assumed to be related to the very low strength of aliasing artifact on recaptured images from ICL database when compared to ROSE database, so that the forensic detector, which has been trained on the ROSE database and which might use aliasing artifact as the most important clue for classification, will misclassify ICL recaptured images as single captured ones. In addition, in order to well understand the obtained results, we will in the following check the matrix of average correlation coefficients of the horizontal residues from the two databases as given in Eqs. (4) and (5) for ICL database and in Eqs. (6) and (7) for ROSE database.

$$\bar{\mathbf{M}}_{single}^{(ROSE)} = \begin{bmatrix} -0.1186 & 0.0614 & 0.3322 & 0.0705 & -0.1159 \\ -0.2502 & 0.0267 & 0.6549 & 0.0390 & -0.2488 \\ -0.3566 & -0.0565 & 1.0000 & -0.0564 & -0.3565 \\ -0.2489 & 0.0389 & 0.6549 & 0.0268 & -0.2501 \\ -0.1160 & 0.0704 & 0.3322 & 0.0614 & -0.1184 \end{bmatrix}. \tag{6}$$

$$\bar{\mathbf{M}}_{recapture}^{(ROSE)} = \begin{bmatrix} -0.0297 & -0.0199 & 0.4367 & -0.0213 & -0.0336 \\ -0.0934 & -0.0587 & 0.6563 & -0.0623 & -0.0929 \\ -0.1210 & -0.2149 & 1.0000 & -0.2149 & -0.1210 \\ -0.0928 & -0.0622 & 0.6563 & -0.0587 & -0.0934 \\ -0.0335 & -0.0212 & 0.4367 & -0.0200 & -0.0298 \end{bmatrix}. \tag{7}$$

We can observe that CC values of single captured images from the two databases are close to each other, especially for the elements in the second to fourth columns of the matrices (compare elements in Eqs. (4) and (6)), which

explains, to some extent, the relatively high classification accuracy on single captured images even in the very challenging inter-database experiments. However, the alteration pattern between CC values of single captured and recaptured images is quite different on the two databases. More precisely, on the ROSE database, CC values in the second and fourth columns are lower in recaptured images than their counterparts in single captured images (compare these matrix elements in Eqs. (6) and (7)), which is probably induced by the aliasing distortion; meanwhile, elements in the first and fifth columns are higher in recaptured images and we assume that this is due to the blurriness introduced by image recapture. Concerning the alteration pattern of CC values on ICL database, as analyzed in Section 3.2.2, CC values from the second to fourth columns are higher in recaptured images mainly due to the dominating blurriness alteration, and CC values at the four corners are lower in recaptured images which might be induced by the slight aliasing artifact. We assume that this difference in alteration pattern of CC values is due to the different characteristics of introduced distortion by quite different image recapturing techniques used to construct the two databases, in particular the different strength of aliasing artifact. This can, to some extent, explain the low classification accuracy on recaptured images in the inter-database setting.

It is worth mentioning that existing image recapture forensic methods also fail under this very challenging inter-database setting. For example, the method of Thongkamwitoon et al. (2015), when trained on ICL database and tested on ROSE database, gives 66.95%, 26.89% and 46.67% respectively for accuracies on single captured, recaptured and all images (all lower than the results of our method shown in the second row of Table 6). As mentioned above, we believe that the encountered difficulty is mainly related to the different statistical properties of recaptured images on the two databases and in a broader sense related to the problem of a "mismatch" between the distribution of the training data and that of the testing data, a common and open research problem for image forensics and for machine learning in general. This problem will be briefly discussed in the last paragraph of this subsection.

As shown and discussed earlier, our method has poor performance under the inter-database setting because of a "mismatch" between data used for training and those used for testing. A straightforward solution is to let the forensic detector see, during its training, some data similar to those that it will encounter during testing. This is actually the motivation of conducting experiments under the combined-database setting in which we use both ICL and ROSE images for training and testing. In order to control the total time required for 10 runs of experiments, for each run we randomly select 50% of the images in the combined data set for training and the remaining images for testing. The last row of Table 6 presents the classification accuracies under this combined-database setting. All the three accuracies, respectively on single captured, recaptured and all images, are very high. This implies that the proposed image statistics feature is discriminative not only on the individual ICL and ROSE database which have different kinds of recapture distortions, but also very effective when we mix the images in two databases together. In other words, the (loose) assumption and basic idea of our method, *i.e.*, image recapturing introduces subtle yet detectable alteration of local image statistics, are valid for images from different sources and obtained with different recapturing techniques. More precisely, there is nearly no overlapping in the feature space for the two kinds of images (*i.e.*, single captured or recaptured) even we mix images from different databases, as reflected by the very high classification accuracies given in the last row of Table 6. This demonstrates the high discriminability of the proposed statistics feature. In addition, like many machine-learning-based algorithms, including more data from different sources during the training stage is often beneficial to find an accurate boundary between the decision regions here corresponding respectively to single captured and recaptured images.

At last, we would like to point out that in fact an elegant and potentially effective way to obtain improved performance under the inter-database setting is to derive and use an appropriate *domain adaptation* technique, either in a supervised or in an unsupervised setting. Domain adaptation (Ben-David et al., 2010) is an active research field in machine learning to solve problems

encountered when applying a well trained model to perform classification of a set of testing data which follow a different distribution from that of the data used for training of the model. The resolution of such problems in recapture forensics, or in the field of image forensics in general, is out of the scope of this manuscript and we would like to leave this interesting and open question, yet having received little attention among the multimedia security community, as an important part of future work.

4.4. Discussion

To our knowledge, our method is the first in the literature that reports consistently good forensic results on the two existing popular and large-scale databases of high-resolution and high-quality recaptured images. The proposed image statistics feature appears to be discriminative enough to be able to well detect recaptured images of different properties, *e.g.*, with or without visible aliasing-like distortions as comprised in the two databases.

If we take a close look at the proposed feature, we will have the intuition that it is closely related to the multivariate Gaussian model for local image statistics in the pixel differential domain. In fact, a multivariate Gaussian is completely determined by its mean vector and co-variance matrix. As argued and observed in many references, *e.g.*, in (Hyvärinen et al., 2009), the pixel means in a differential domain are very close to zero, therefore the co-variance matrix is the more informative parameter of the multivariate Gaussian model when using it to describe local image statistics. What we include in our feature vector is not exactly the co-variance matrix, but is closely related to it and can be considered as a simplified version of it. More precisely, if we compute all the pair-wise co-variances within 5×5 patches, we will obtain a 25×25 matrix, and to some extent we can say that our feature vector only comprises 14 elements of this matrix (after proper normalization). This simplification is reasonable because of the redundancy in the co-variance matrix and some intrinsic properties of local image statistics (*e.g.*, the approximate translation invariance). In all, loosely speaking, our forensic method makes an implicit

assumption of multivariate Gaussian model for the pixel values' statistics in the residue images. Following this research line, it would be interesting to attempt to use more sophisticated and more accurate (non-Gaussian) image statistics models, such as those described in (Hyvärinen et al., 2009), for the purpose of image recapture forensics.

Another promising research direction, also closely related to image statistics, is the derivation of a kind of optimal filter(s) used for transforming to image differential domain(s). We tried in this work a set of different filters for being used in Eq. (1) and found that the simple filters in Eq. (2) gave the best results among the tested candidates (also see discussion in the paragraph below Eq. (2)). In the literature of natural image statistics research, such filters have been optimized for maximizing the variances or sparseness of the filtered results (Hyvärinen et al., 2009), for example in an attempt to model the functional mechanism of visual cells in human brain. One interesting question would be the following: Would it be possible to find one or several optimal filters that lead to optimal classification accuracy of image recapture forensics? Studies on this question, along with the use of more sophisticated image statistics models as described above, appear to be a very promising future working direction.

5. Conclusion

In this manuscript, we presented an image-statistics-based method for detecting recaptured images. Different from the current trend of developing more and more sophisticated recapture forensic methods by either combining multiple features or using complex machine learning tools, our method is conceptually simple and computationally efficient. More precisely, we revisited the idea of using image statistics for recapture forensics and proposed very simple local image statistics of second-order mixed moments in pixel differential domains. The feature extraction procedure is extremely simple and very fast, and the corresponding forensic detector can successfully accomplish the task of distinguishing between single captured and recaptured images on two large databases compris-

ing high-resolution and high-quality recaptured images from LCD screens. Our method achieves slightly higher classification accuracy when compared with state-of-the-art methods, while remaining very simple, fast and stable.

Concerning future work, as mentioned in Section 4.4, we would like to use more advanced and more accurate image statistical models and to optimize the coefficients in filter kernels for better image recapture forensics. Another interesting future working direction is to use the proposed recaptured image detection algorithm in closely related practical applications. In particular, it would be interesting to integrate our recapture forensic method in a biometric authentication system as an important part of the spoofing detection module (Menotti et al., 2015), *e.g.*, for detecting spoofing attacks in which an attacker attempts to deceive an iris, face or fingerprint authentication system by showing to the system a recaptured image of biometric trait of an authorized person. Finally, we plan to study the domain adaptation technique in order to facilitate the practical deployment of image recapture forensic detectors.

Acknowledgments

The author would like to thank the anonymous reviewers for their valuable suggestions and comments which greatly improve the quality of this manuscript. We would like to thank Dr. T. Thongkamwitoon for making his recaptured images data set and the source code of his image recapture detection method freely available on-line, as well as for some email exchanges. The research in this manuscript also used the ROSE Recaptured Image Database made available by the ROSE lab at the Nanyang Technological University, Singapore. This work has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program "Investissement d'avenir".

References

Abdi, H., 2007. Multiple correlation coefficient, in: Salkind, N. (Ed.), Encyclopedia of Measurement and Statistics. SAGE Publications, pp. 648–651.

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J., 2010. A theory of learning from different domains. Machine Learning 79, 151–175.

Cao, H., 2010. Statistical Image Source Model Identification and Forgery Detection. Ph.D. thesis. Nanyang Technological University.

Cao, H., Kot, A.C., 2010. Identification of recaptured photographs on LCD screens, in: Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1790–1793.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–297.

Fan, Z., de Queiroz, R., 2003. Identification of bitmap compression history: JPEG detection and quantizer estimation. IEEE Transactions on Image Processing 12, 230–235.

Gao, X., Ng, T.T., Qiu, B., Chang, S.F., 2010. Single-view recaptured image detection based on physics-based features, in: Proc. of IEEE International Conference on Multimedia & Expo, pp. 1469–1474.

Heikkilä, M., Pietikäinen, M., Schmid, C., 2009. Description of interest regions with local binary patterns. Pattern Recognition 42, 425–436.

Hyvärinen, A., Hurri, J., Hoyer, P.O., 2009. Natural Image Statistics – A probabilistic approach to early computational vision. Springer.

Ke, Y., Shan, Q., Qin, F., Min, W., 2013. Image recapture detection using multiple features. International Journal of Multimedia and Ubiquitous Engineering 8, 71–82.

Li, B., Shi, Y.Q., Huang, J., 2008. Detecting doubly compressed JPEG images by using mode based first digit features, in: Proc. of IEEE International Workshop on Multimedia Signal Processing, pp. 730–735.

Li, R., Ni, R., Zhao, Y., 2015. An effective detection method based on physical traits of recaptured images on LCD screens, in: Proc. of International Workshop on Digital-Forensics and Watermarking, pp. 107–116.

Lyu, S., Farid, H., 2005. How realistic is photorealistic? IEEE Transactions on Signal Processing 53, 845–850.

Mahdian, B., Novozamsky, A., Saic, S., 2015. Detecting cyclostationarity in re-captured LCD screens. Journal of Forensic Research 6, 294:1–294:6.

Menotti, D., Chiachia, G., Pinto, A., Schwartz, W.R., Pedrini, H., Falcão, A.X., Rocha, A., 2015. Deep representations for iris, face, and fingerprint spoofing detection. IEEE Transactions on Information Forensics and Security 10, 864–879.

Ng, T.T., Chang, S.F., 2013. Discrimination of computer synthesized or recaptured images from real images, in: Sencar, H.T., Memon, N. (Eds.), Digital Image Forensics. Springer-Verlag, pp. 275–309.

Ni, R., Zhao, Y., Zhai, X., 2015. Recaptured images forensics based on color moments and DCT coefficients features. Journal of Information Hiding and Multimedia Signal Processing 6, 323–332.

Ojala, T., Pietikäinen, M., Mäenpää, P., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 971–987.

Piva, A., 2013. An overview on image forensics. ISRN Signal Processing 496701, 1–22.

Stamm, M.C., Liu, K.J.R., 2010. Forensic detection of image manipulation using statistical intrinsic fingerprints. IEEE Transactions on Information Forensics and Security 5, 492–506.

Stamm, M.C., Wu, M., Liu, K.J.R., 2013. Information forensics: An overview of the first decade. IEEE Access 1, 167–200.

Thongkamwitoon, T., Muammar, H., Dragotti, P.L., 2015. An image recapture detection algorithm based on learning dictionaries of edge profiles. IEEE Transactions on Information Forensics and Security 10, 953–968.

Yin, J., Fang, Y., 2012. Digital image forensics for photographic copying, in: Proc. of SPIE-IS&T Electronic Imaging, pp. 83030F:1–83030F:7.