

Article

Hybrid Image-Retrieval Method for Image-Splicing Validation

Nam Thanh Pham ¹, Jong-Weon Lee ² , Goo-Rak Kwon ³ and Chun-Su Park ^{4,*}

¹ Department of Digital Contents, Sejong University, Seoul 05006, Korea; pham@sju.ac.kr

² Department of Software, Sejong University, Seoul 05006, Korea; jwlee@sejong.ac.kr

³ Department of Information and Communication Engineering, Chosun University, Gwangju 61452, Korea; grkwon@chosun.ac.kr

⁴ Department of Computer Education, Sungkyunkwan University, Seoul 03063, Korea

* Correspondence: cspk@skku.edu

Received: 22 November 2018; Accepted: 10 January 2019; Published: 14 January 2019



Abstract: Recently, the task of validating the authenticity of images and the localization of tampered regions has been actively studied. In this paper, we go one step further by providing solid evidence for image manipulation. If a certain image is proved to be the spliced image, we try to retrieve the original authentic images that were used to generate the spliced image. Especially for the image retrieval of spliced images, we propose a hybrid image-retrieval method exploiting Zernike moment and Scale Invariant Feature Transform (SIFT) features. Due to the symmetry and antisymmetry properties of the Zernike moment, the scaling invariant property of SIFT and their common rotation invariant property, the proposed hybrid image-retrieval method is efficient in matching regions with different manipulation operations. Our simulation shows that the proposed method significantly increases the retrieval accuracy of the spliced images.

Keywords: image splicing; localization; image retrieval; Zernike moment; SIFT

1. Introduction

Digital images, portable cameras, and photo-editing software have become increasingly popular in the past decades. This popularization has enabled the easy manipulation of digital images even by unpracticed users. Nowadays, image splicing, which copies one or more regions from the source image and pastes them to the destination image, is one of the most popular image-manipulation methods.

In the image-splicing scenario, a spliced region might not be exactly the same as the original region since it usually undergoes a sequence of postprocessing operations such as rotation, scaling, edge softening, blurring, denoising, and smoothing for a better visual appearance. Therefore, human beings may easily be deceived by spliced images.

In the literature, a large number of algorithms have been introduced to effectively detect image splicing [1–6] and some algorithms achieved nearly perfect detection performance [2,3]. In recent years, researchers have been focusing more on image-splicing localization and achieved promising results [7] thanks to the advances in machine learning and deep learning [8–10]. Tampered regions of grayscale images were localized by different image-authentication techniques [11,12]. In Reference [6], a convolutional neural network (CNN)-based algorithm was proposed to extract features capturing characteristic traces from different camera models. These features were then utilized as the input for an iterative clustering algorithm to estimate the tampering mask. The difference between noise levels of tampered and original regions was employed to find the splicing traces [8,13]. The noise level was estimated using principal component analysis and then clustered by a *k*-means algorithm to localize the spliced regions [13]. Nonlinear camera response function was individually used in Reference [14]

or combined with noise-level function to exploit their strong relationship to localize the forged edges using a CNN [8]. Gamma transformation was used to detect splicing forgery whilst spliced region was localized based on the probabilities of overlapping blocks and pixels of the input image being gamma-transformed [15]. In Reference [16], spatial structure on the boundary between spliced and background regions was trained to localize splicing, copy-move, and removal manipulations.

In contrast to those abovementioned blind-splicing localization methods, context-based search-and-compare approaches were proposed in References [17,18] to localize spliced regions. Specifically, a spliced image was used as a query image in the problem of image retrieval among the database of authentic images. Finally, the retrieved result then combined with the corresponding query image to find the difference, which is the localization result. A novel deep CNN was developed to compute the probability that one image had been spliced to another image, and then splicing masks were localized in both images [9]. These approaches achieved high localization performance in comparison with blind methods, but they can only be applied if the original authentic image dataset is given.

In this paper, we go one step further by providing solid evidence for image manipulation. If a certain image is proved to be the spliced image, the proposed method retrieves the original authentic images that were used to generate the spliced image to yield a clue about the spliced images. First of all, the proposed algorithm localizes tampered regions in the spliced image by using one of the state-of-the-art localization algorithms [19]. Then, the proposed algorithm retrieves the original authentic images using a Hamming Embedding (HE) encoding-based retrieval algorithm with Zernike [20] and Scale Invariant Feature Transform (SIFT) [21] descriptors. In the literature, a few studies were proposed to find the provenance of manipulated images. In Reference [22], the provenance of spliced images was searched in two tiers using Speeded-Up Robust Features. The first tier was responsible to find the host (target) images, and then the search was refined in the second tier to find the donor (source) images. Provenance graph-construction methods were proposed in References [23,24] to show the relationships among images with multiple manipulation operations, such as splicing, spatial operations and enhancement. The experimental results in References [22–24] were performed using only a small number of test images and the results show that their methods achieved moderate retrieval performance for the source images. This paper aims to address the shortcomings of previous works and provide a convincing provenance of spliced images. The rest of this paper is organized as follows. Section 2 describes Zernike moment features and SIFT features, including their applications in previous works. Then, the image-splicing localization method is presented in Section 3 and the proposed retrieval algorithm is introduced in Section 4. In Section 5, we provide the comparative experimental results of the proposed method. Finally, our conclusions are drawn in Section 6.

2. Related Work

In this Section, we give a brief overview of Zernike moment features [20,25] and SIFT features [21]. Given a continuous image function $f(x, y)$, Zernike moments are the projection of the image function onto a set of basis functions called Zernike polynomials. This set of complex polynomials is complete and orthogonal over unit circle $x^2 + y^2 = 1$. Let us denote these polynomials as $V_m^n(x, y)$, suppose that m is a non-negative integer, and n is an integer satisfies that $m - |n|$ is a non-negative even number. Zernike polynomials are defined as follows:

$$\begin{aligned} V_m^n(x, y) &= V_m^n(\rho, \theta) \\ &= R_m^n(\rho) \exp(jn\theta), \end{aligned} \quad (1)$$

where (ρ, θ) represents the polar co-ordinates of (x, y) , j is an imaginary unit number, and the radial polynomial $R_m^n(\rho)$ is given by:

$$R_m^n(\rho) = \sum_{s=0}^{\frac{m-|n|}{2}} \frac{(-1)^s (m-s)! \rho^{m-2s}}{s! \left(\frac{m+|n|}{2} - s\right)! \left(\frac{m-|n|}{2} - s\right)!}. \quad (2)$$

Note that $R_m^n = R_m^{-n}$. The Zernike moment with order m and repetition n for image $f(x, y)$ that vanishes outside the unit circle is defined as:

$$A_m^n = \frac{m+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x, y) \overline{V_m^n(\rho, \theta)} dx dy, \quad (3)$$

where \bar{z} denotes the complex conjugate of z . It is clear that, since $A_m^n = \overline{A_m^{-n}}$, then $|A_m^n| = |A_m^{-n}|$. The symmetric and antisymmetric properties of Zernike moments help reduce their computational complexity [26]. We replace integrals in Equation (3) with summations to obtain the formula for digital image $I(x, y)$ as follows:

$$A_m^n = \frac{m+1}{\pi} \sum_x \sum_y I(x, y) \overline{V_m^n(\rho, \theta)}, \quad (4)$$

where $x^2 + y^2 \leq 1$. The magnitudes of the Zernike moments, $|A_m^n|$, are used as the image features. According to References [20,25–29], these values are invariant against rotation; hence, Zernike moment features can handle well rotated spliced regions.

The SIFT keypoints are invariant to image rotation and scaling, and robust for noise and illumination changes [21]. Such keypoints of input image $I(x, y)$ are detected by finding the local extrema of $D(x, y, \sigma)$, which is the convolution of that image with a difference of Gaussian between adjacent image scales $k\sigma$ and σ :

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y), \quad (5)$$

where $*$ denotes the convolution operation, k is the constant multiplicative factor, and Gaussian function $G(x, y, \sigma)$ at scale σ is calculated as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (6)$$

As shown in Figure 1, the red pixel is marked as the keypoint if it is greater or smaller than all of its 26 neighbors, including 8 neighbors in the same scale (yellow pixels) and 9 neighbors in the corresponding positions of the scales above and below (green pixels). The corresponding descriptors of all the keypoints are extracted by computing their gradient magnitudes and orientations to form the features [21].

Zernike moments have a wide range of applications in different fields such as image recognition and classification [20,25,27,28], copy-move-forgery detection [29–33], video-forgery detection [34], watermark detection [35–37], and medical-image retrieval [38]. In the copy-move forgery-detection problem, Zernike moment features-based methods notably showed impressive performances to different kinds of transformations in comparison with other approaches [29–31]. On the other hand, SIFT features were used extensively in image-retrieval studies [39–43] and copy-move-forgery detection [32,44,45]. A coupled multi-index method was proposed in Reference [46] to exploit the feature fusion of local color feature and SIFT feature where both kinds of descriptors are extracted for all keypoints. In Reference [47], SIFT descriptors were quantized to descriptive and discriminative bit-vectors, called binary SIFT, to avoid high computation cost. Further, SIFT-based methods, CNN-based methods, and their connections were presented in Reference [10] to give a comprehensive overview of recent image-retrieval studies.

However, in the image-splicing scenario, the straightforward application of the Zernike moment features or the SIFT features is not efficient to retrieve spliced images because of the wide variety of shapes, sizes, and textures of the spliced regions. Therefore, in this paper, we propose a hybrid method that combines two advanced feature types, the Zernike moment features and the SIFT features, for solving the image-splicing retrieval problem.

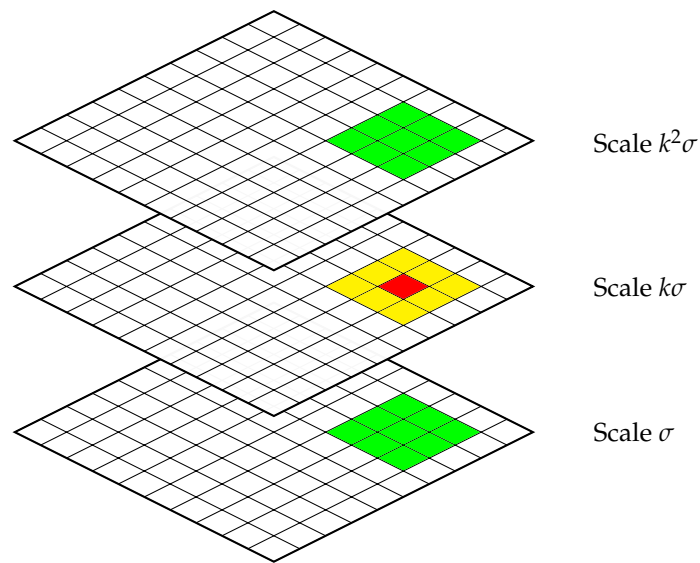


Figure 1. Detecting SIFT keypoints in one scale by finding local maxima or minima.

3. Image-Splicing Localization

Suppose that a spliced image $I(x, y)$ is composed of two authentic images, $S(x, y)$ and $T(x, y)$, which are the source and target images, respectively. Here, one or more regions of the source image can be copied and pasted to the target image to create the spliced image. Afterward, this resultant image $I(x, y)$ is usually edited by some postprocessing operations to make it look realistic and rational. Given a set of spliced images, this study aims to retrieve their original-source images and target images. In Sections 3 and 4, our proposed approach is presented to solve this problem. In Figure 2, we depict the overall framework of the proposed method, which consists of two main stages: image-splicing localization and image retrieval. In the first stage, we firstly segment the input spliced image $I(x, y)$ into the spliced regions (foreground) $F(x, y)$ and the background region $B(x, y)$. The detailed splicing localization algorithm is described in Section 3. The segmented regions are subsequently utilized as the input of the second stage. The proposed image-retrieval method is presented in Section 4, where we query $F(x, y)$ and $B(x, y)$ to find $S(x, y)$ and $T(x, y)$, respectively.

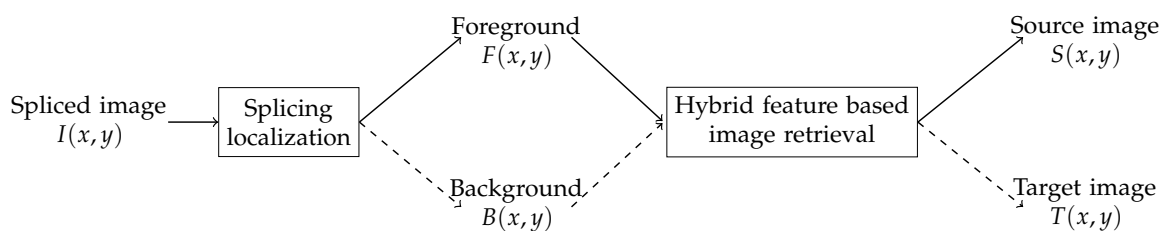


Figure 2. General framework of the proposed method for image-splicing retrieval.

In order to retrieve all images that were used to compose the spliced image, we need to segment the background and spliced regions of that spliced image to separately extract features. In this paper, we adopted a fast and efficient splicing-localization algorithm introduced in Reference [19]. In their proposed algorithm, the input image is firstly converted to YUV space, and each color component of the image is divided into nonoverlapping 8×8 blocks. These blocks of three color components are then transformed to the Discrete Cosine Transform (DCT) domain. The histograms of DCT coefficients of all blocks in different color channels are subsequently used to construct a block posterior probability map. Afterward, a Support Vector Machine classifier utilizes this probability map as the feature to classify whether the 8×8 block is tampering or not. Note that we carry out some further postprocessing operations, including median filtering, morphological erosion, and dilation, to remove noises and

to bridge the gaps of the generated tampering map to localize the spliced regions. This tampering image is detected in block level; hence, it is then upscaled to obtain the final localized splicing image. To quantitatively evaluate the performance of splicing localization, we adopted two metrics for spliced regions of a detected spliced image [31,44], precision M_P and recall M_R , which are defined as follows:

$$M_P = \frac{\text{\#correctly detected pixels}}{\text{\#all detected pixels}}, \quad (7)$$

and

$$M_R = \frac{\text{\#correctly detected pixels}}{\text{\#all spliced pixels}}. \quad (8)$$

There is a trade-off between precision and recall; consequently, to consider both these measures, we computed their harmonic mean M_F , called F_1 -score, as follows:

$$M_F = \frac{2M_P M_R}{M_P + M_R}. \quad (9)$$

In the proposed method, the localization of spliced regions is considered as a correct localization if $M_F > 0.7$. These correctly localized spliced regions and the corresponding background of the same spliced images are then used for retrieval in the next stage. Examples of spliced images, segmentation of spliced and background regions, and ground truth images are presented in Figures 3 and 4.

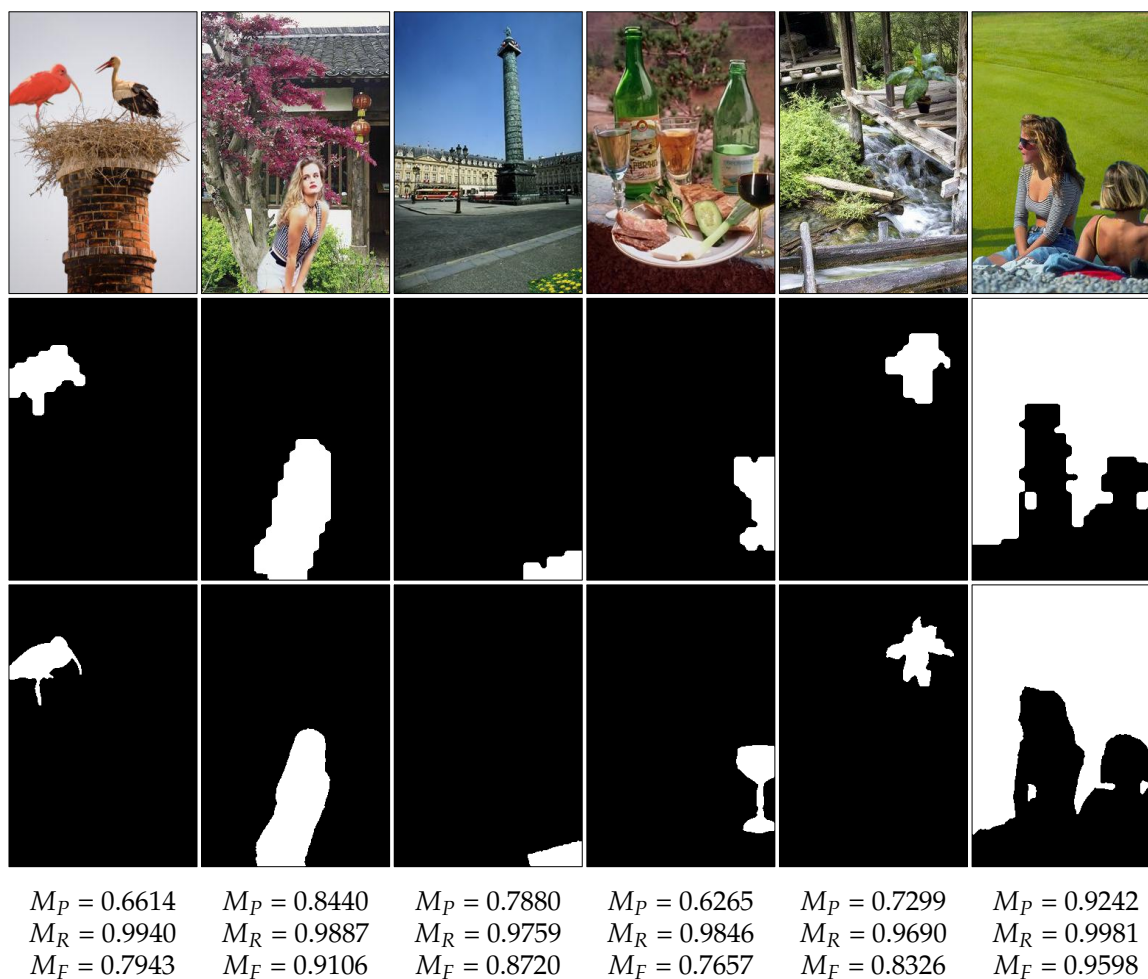


Figure 3. Image-splicing localization example. Spliced images are shown in the **top row**, and their corresponding spliced masks and ground truth images are presented in the **second row** and the **third row**, respectively. Finally, precision, recall, and F_1 -score of each localized splicing image are given.

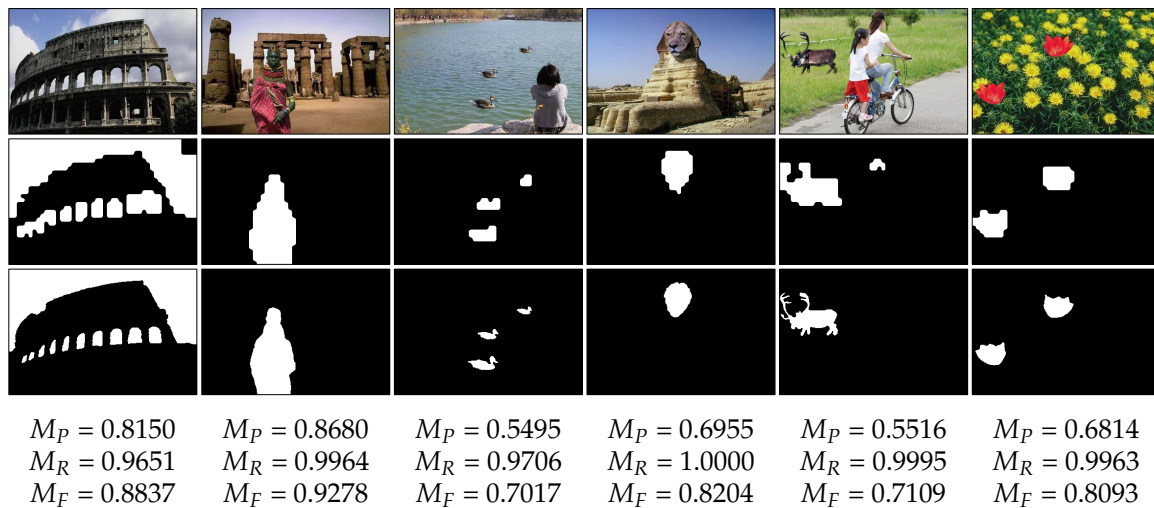


Figure 4. Image-splicing localization example. Spliced images are shown in the **top row**, and their corresponding spliced masks and ground truth images are presented in the **second row** and the **third row**, respectively. Finally, precision, recall, and F_1 -score of each localized splicing image are given.

4. Image-Splicing Retrieval

Rotation and scaling are two operations that affect the shape and direction of spliced regions. Therefore, we take advantage of the rotation invariant property of the Zernike moment features and the rotation scale invariant property of the SIFT features to trace spliced regions. Whereas the SIFT features-based method (the SIFT method) effectively extracts features in textured regions, it usually fails in smooth regions. The Zernike moment features-based method (the Zernike method), in contrast, successfully addressed the shortcomings of the SIFT method. In addition, the Zernike moment features extraction is a block-based method, while SIFT features extraction is a keypoint-based method. For those reasons, in the proposed method, the Zernike moment features and the SIFT features are independently extracted to handle different situations in the spliced-image retrieval problem.

4.1. Bag-of-Features-Based Image-Retrieval Using Hamming Embedding

In image retrieval, images are represented by descriptive features, which are used to evaluate similarity or dissimilarity between a pair of images. Since the splicing operation may rotate, scale, translate the forged regions, the extracted features of the images should be invariant to these transformations. The features generated from Zernike moments and SIFT have such noteworthy characteristics. In this Section, we present the HE encoding [40,48] and bag-of-features-based [49–51] image-retrieval method. Suppose that query region Q is described by a set of n local descriptors, $X_Q = \{x_1^Q, x_2^Q, \dots, x_n^Q\}$. All these descriptors are mapped into a visual vocabulary set $W = \{w_1, w_2, \dots, w_k\}$ by a k -mean vector quantizer q . In other words, q maps x_i^Q to a visual word where $q(x_i^Q) \in W$. It is worthwhile to note that, in the proposed design, both the spliced region and the background region could be the query region. We define a set of index of descriptors in X_Q that are assigned to a particular visual word w as $I_w^Q = \{i \mid q(x_i^Q) = w\}$. In order to match different descriptors, each descriptor is represented as a binary signature [40]. We define d as the number of bits used for descriptor representation, i.e., $b(x_i^Q) = (b_1(x_i^Q), b_2(x_i^Q), \dots, b_d(x_i^Q))$ is a binary signature of descriptor x_i^Q . The Hamming distance between the binary signatures of two descriptors, u and v , is computed as follows:

$$h(b(u), b(v)) = \sum_{i=1}^d |b_i(u) - b_i(v)|. \quad (10)$$

Let us denote $M_w(X_Q, X_D)$ as a function between two sets of descriptors X_Q and X_D assigned to the same visual word w as:

$$M_w(X_Q, X_D) = \sum_{i \in I_w^Q} \sum_{j \in I_w^D} f\left(h\left(b(x_i^Q), b(x_j^D)\right)\right), \quad (11)$$

where weighting function f is calculated as a Gaussian function [40]:

$$f(h) = \begin{cases} e^{-16h^2/d^2}, & \text{if } h \leq 3d/8, \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

Finally, the similarity between two images, Q and D , is defined as:

$$\text{sim}(Q, D) = \frac{\sum_{w \in W} c_w M_w(X_Q, X_D)}{\sqrt{\left(\sum_{w \in W} c_w M_w(X_Q, X_Q)\right) \left(\sum_{w \in W} c_w M_w(X_D, X_D)\right)}}, \quad (13)$$

where constant c_w is the inverse document frequency [52] of visual word w in W .

4.2. Hybrid Features-Based Image Retrieval

In this paper, we propose a hybrid image retrieval method exploiting Zernike and SIFT features. In order to effectively combine these two methods in the retrieval process, we evaluate their retrieval performance separately. Figure 5 illustrates clusters of spliced regions, which are correctly retrieved by either the Zernike method, the SIFT method or both methods. Two measurements of spliced regions, the extent and the smoothness, are represented by the number of pixels and the number of SIFT keypoints, respectively. The Zernike method correctly retrieves the higher number of images in comparison with the SIFT method. In addition, for small and smooth regions, the former also shows a better performance than the latter. Therefore, at first, the proposed method performs image retrieval using the Zernike moment features of the query regions. Subsequently, based on the intermediate retrieval results, the proposed method decides whether the SIFT features need to be utilized to improve the retrieval performance.

We introduce a method to effectively utilize two kinds of features in image retrieval. Firstly, we apply the localization algorithm [19] to the spliced image $I(x, y)$ to segment foreground (spliced region) and background, say $F(x, y)$ and $B(x, y)$, respectively. These regions are sequentially used as the queries for image retrieval process. Assume that the current query region is $F(x, y)$. Then, the proposed method extracts the Zernike moment features of $F(x, y)$ and evaluates the certainty score of the top-hit retrieval image using extracted features. We define s as the certainty score in Equation (15). If the score is positive, the top-hit retrieval image is considered as the source image $S(x, y)$, which was used for image splicing. In contrast, if the certainty score s is smaller than or equal to 0, the SIFT features are subsequently used to retrieve the original source image of spliced region $F(x, y)$. Figure 6 shows the pipeline of the proposed spliced image retrieval method for $F(x, y)$. Similarly, if the background region $B(x, y)$ is queried, the target image of image splicing, $T(x, y)$, is retrieved.

In the first stage of the proposed image-retrieval process using the Zernike moment features, each retrieved image has a score that represents the similarity of that image with the input image. Suppose that I_1 and I_2 are the first and second top-hit retrieval images of the input image with voting scores s_1 and s_2 , respectively. It is worth pointing out that, in the spliced image retrieval problem, the top-hit retrieved image should be the only correct result for each query. We consider I_1 as the correctly retrieved image if the ratio s_1/s_2 is greater than a specific value in different intervals of s_1 .

In contrast, the SIFT features are taken into account in the second stage of the proposed method if one of the following conditions happens:

$$\begin{cases} s_1 < \alpha_1, \\ \alpha_1 < s_1 < \alpha_2 & \text{and } s_1/s_2 < \lambda_1, \\ \vdots \\ \alpha_N < s_1 < \alpha_{N+1} & \text{and } s_1/s_2 < \lambda_N, \end{cases} \quad (14)$$

where we use α_i ($i = 1, 2, \dots, N$) as bounds of different intervals for top-hit score s_1 and λ_i ($i = 1, 2, \dots, N$) as thresholds for s_1/s_2 in each corresponding interval of s_1 . Since s_1 is a positive integer, in the experiments, we avoid setting α_i ($i = 1, 2, \dots, N$) as integer numbers. We define s as the certainty score for the top-hit image in the first retrieval stage, which is calculated based on Equation (14) as follows:

$$s = \sum_{i=0}^N \left[e^{-(s_1 - \alpha_i)(s_1 - \alpha_{i+1})} \right] (s_1 - \lambda_i s_2), \quad (15)$$

where we define 2 more parameters for this general formula, $\lambda_0 = 0$ and $\alpha_0 \in ([\alpha_1], \alpha_1)$. If s is positive, I_1 is considered as the correct retrieval of the query image region; otherwise, the proposed method does the retrieval process again using SIFT features.

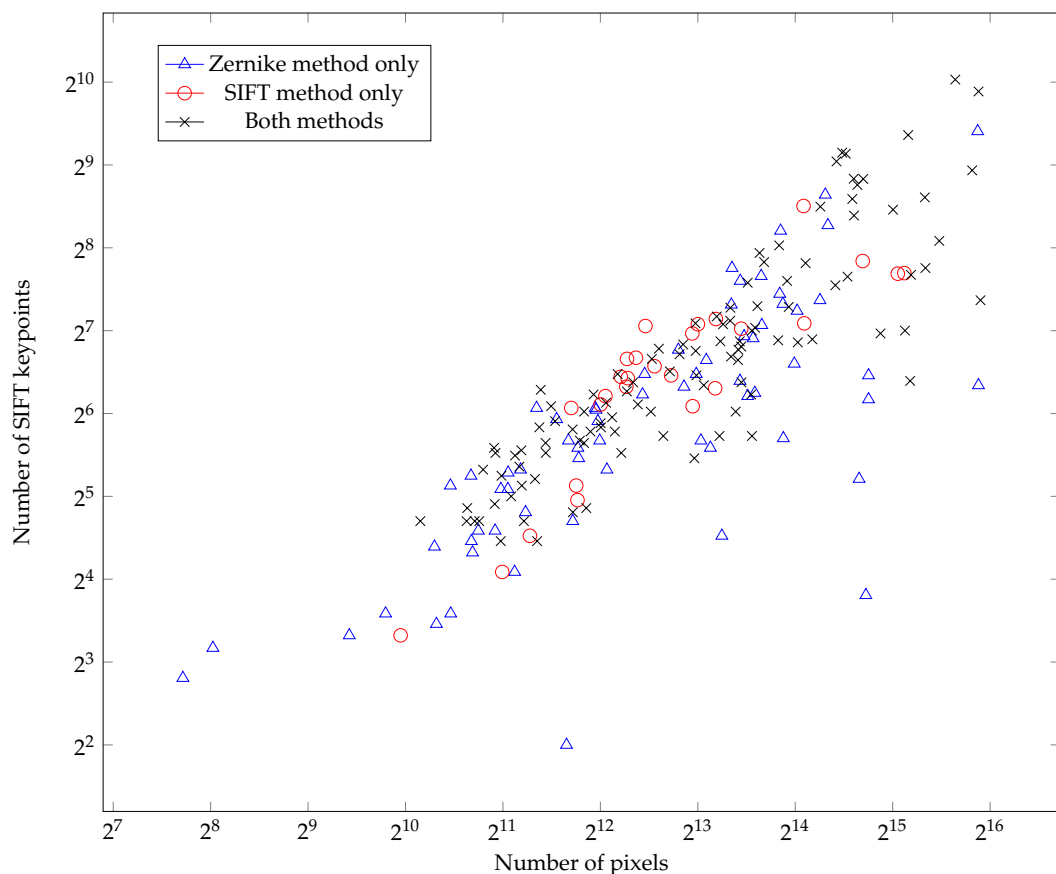


Figure 5. Illustration of spliced regions that are correctly retrieved by Zernike-based, SIFT-based, or both methods. The number of pixels of spliced regions and the number of SIFT keypoints in the corresponding regions are presented.

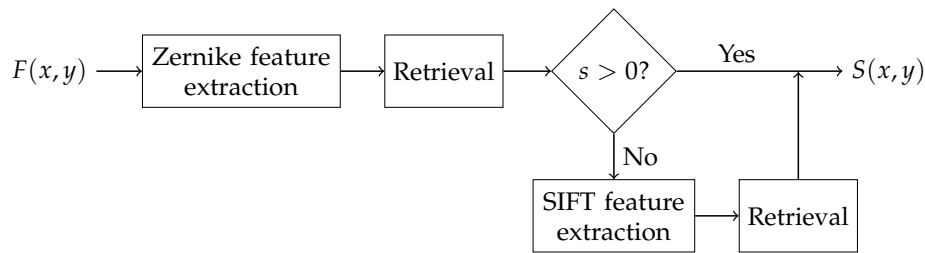


Figure 6. Hybrid features method for spliced image retrieval.

5. Experimental Evaluation

5.1. Dataset and Experimental Setup

There exist several benchmarking datasets for evaluating the performance of image-splicing algorithms. In our simulations, we used CASIA v2.0, a realistic and challenging dataset introduced in Reference [53]. In this dataset, there are 7491 authentic, 1849 spliced, and 3274 copy-move color images with various sizes from 240×160 to 900×600 pixels. However, due to the lack of ground truth images of this dataset, we generate the ground truth images for spliced regions for the evaluation of localization algorithm (Equations (7)–(9)). The ground truth dataset was contributed to the research community, available online at <https://github.com/namtpham/casia2groundtruth>. In the retrieval testing set, we randomly chose 225 spliced images whose spliced regions are correctly localized in all the categories defined in the dataset, such as animal, architecture, character, plant, nature, indoor, and texture. The diversity of characteristics of spliced regions is depicted in Figure 5. Finally, 225 query images, which are equivalent to 450 query regions, are retrieved in 7491 database authentic images. We trained 100,000 visual words for the experiments.

The parameters in Equation (15) were empirically set as presented in Table 1. We carried out the simulations by independently querying spliced regions and background regions. In each case, we performed the image retrieval of the Zernike method and the SIFT method, separately, to compare with the proposed method.

Table 1. Parameters used in the proposed hybrid method.

N	λ_0	λ_1	λ_2	λ_3	λ_4
4	0	1.4	1.25	1.2	1.15
α_0	α_1	α_2	α_3	α_4	α_5
6.4	6.5	11.5	18.5	30.5	1000.5

5.2. Evaluation Measures

In regular image-retrieval problems, a set of images, which are visually similar to the query image, are considered as the correct retrieval results. On the contrary, in the image-splicing retrieval problem of this paper, there is only one correct source image and one correct target image when we query a spliced image. Therefore, we defined a metric quantity $TP_i@N$ as the correctness of the retrieval of i -th query. If the original image of i -th query is retrieved in top N results, then $TP_i@N = 1$; otherwise, $TP_i@N = 0$. The number of correct retrieved images in top N retrieval results is defined as follows:

$$TP@N = \sum_{i=1}^{\#Q} TP_i@N, \quad (16)$$

where $\#Q$ is the total number of queries. The recall at top N retrieval is computed as follows:

$$R@N = \frac{TP@N}{\#Q}. \quad (17)$$

5.3. Image-Splicing Retrieval Results

Table 2 shows the number of query spliced images whose original source and target images are retrieved in the top 1, 10, 50, and 100. As can be seen from the table, the Zernike method performs better than the SIFT method when each of them is independently used for retrieval. The proposed method, a hybrid algorithm of these two features, notably improved the performance of spliced region retrieval and achieved 84.89% accuracy in comparison with 59.56% and 76.44% of the SIFT and the Zernike methods, respectively. Additionally, the proposed method correctly retrieved all the background regions.

Table 2. Spliced-image retrieval results.

Query Region	Features	TP@1	TP@10	TP@50	TP@100	R@1
Foreground (Spliced region)	Zernike	172	183	201	203	76.44%
	SIFT	134	170	184	189	59.56%
	Hybrid	191	200	213	218	84.89%
Background	Zernike	222	225	225	225	98.67%
	SIFT	221	223	225	225	98.22%
	Hybrid	225	225	225	225	100%

The performance of the image-retrieval methods in terms of $R@N$ for background and spliced query regions is given in Figure 7.

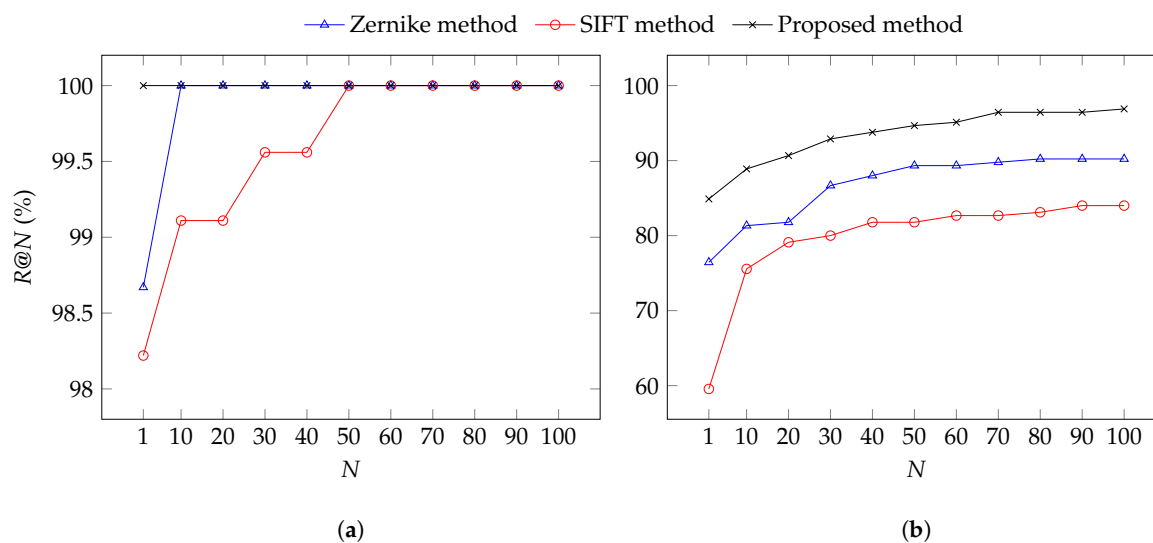


Figure 7. $R@N$ values of image-retrieval methods for querying (a) background regions; (b) spliced regions.

5.4. Analysis of Top-Hit Retrieved Source Image

Image-retrieval methods succeed in finding target images, whereas their performance of spliced-region retrieval is much lower. To evaluate the efficiency of these retrieval algorithms on spliced regions, we classified 225 spliced images into 5 categories based on the ratio between total area of all spliced regions in one spliced image to the area of that image. These ratios of 225 images ranged from 0.21% to 67.60%, and 84% of the testing images have small spliced regions that account for less than 20% of the whole image in term of area. Figure 8 shows the superiority of the proposed method over the SIFT method and the Zernike method in all the above-mentioned categories by giving the numbers of correctly retrieved spliced regions ($TP@1$). In general, the accuracy of retrieving tiny

spliced regions, whose areas are less than 5% of the corresponding spliced images, is lower than that of bigger spliced regions.

Figure 9 gives four examples of spliced regions retrieval of the three methods, the Zernike method, the SIFT method, and the proposed hybrid method. In the first two cases, both the Zernike method and the SIFT method failed to retrieve the source image. On the other hand, the hybrid method correctly retrieved the images in both cases. All methods correctly retrieved the source image in the third case. In contrast, in the last example, all methods retrieved the wrong source image where the Zernike method found the correct spliced region (the houses) from another authentic image, which was taken of the same scene as the source image but from a different view.

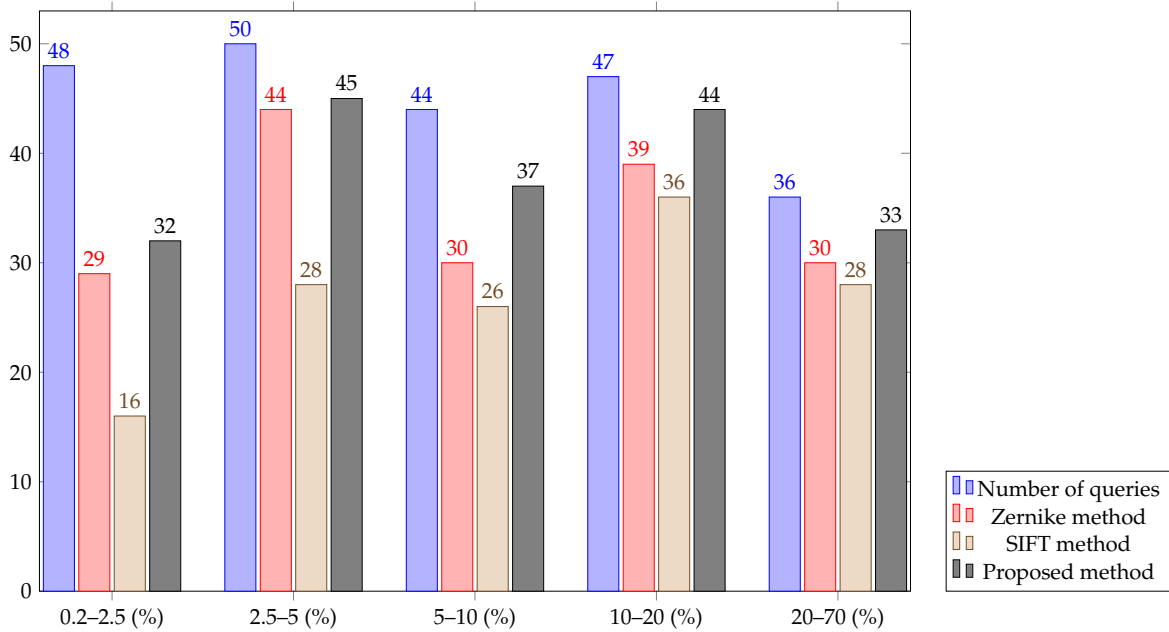


Figure 8. Comparison of number of top-hit retrieved images for querying spliced regions. 225 spliced images are divided into five categories according to the ratios of total area of spliced region in one image to the area of that image. The five categories of the ratios are 0.2%–2.5%, 2.5%–5%, 5%–10%, 10%–20%, and 20%–70%.

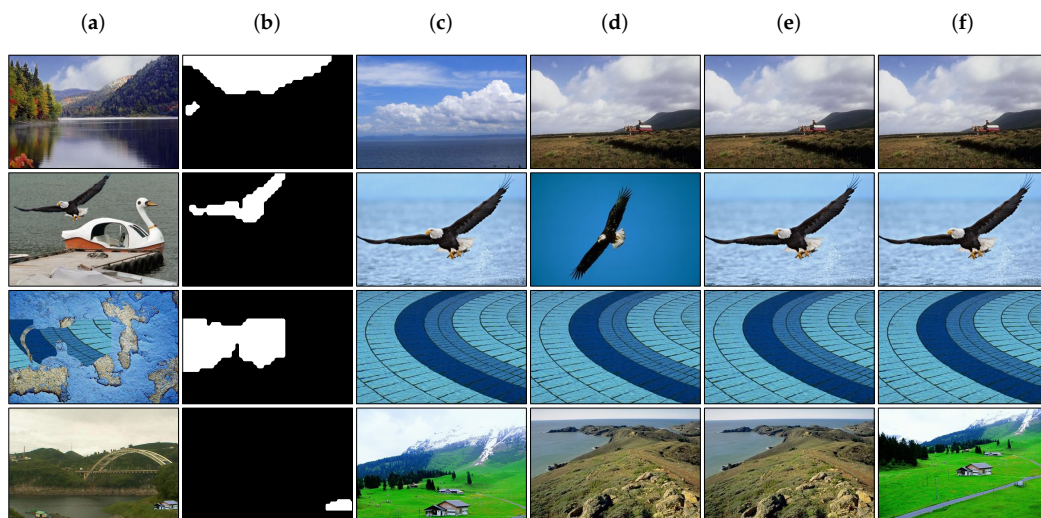


Figure 9. Examples of spliced region retrieval. (a) Spliced images; (b) spliced region masks; (c–e) Top-hit images retrieved for spliced-region queries by the Zernike method, the SIFT method, and the proposed hybrid method, respectively; (f) source images.

5.5. Image-Splicing Validation

An example of image-splicing validation is presented in Figure 10. Two different spliced images were segmented into foreground and background regions. Subsequently, these regions were queried by the hybrid image-retrieval algorithm to find their source and target images. The result shows that they were created by copying different regions in the source image and pasting to the target image. Therefore, the two input images were verified to not be authentic.

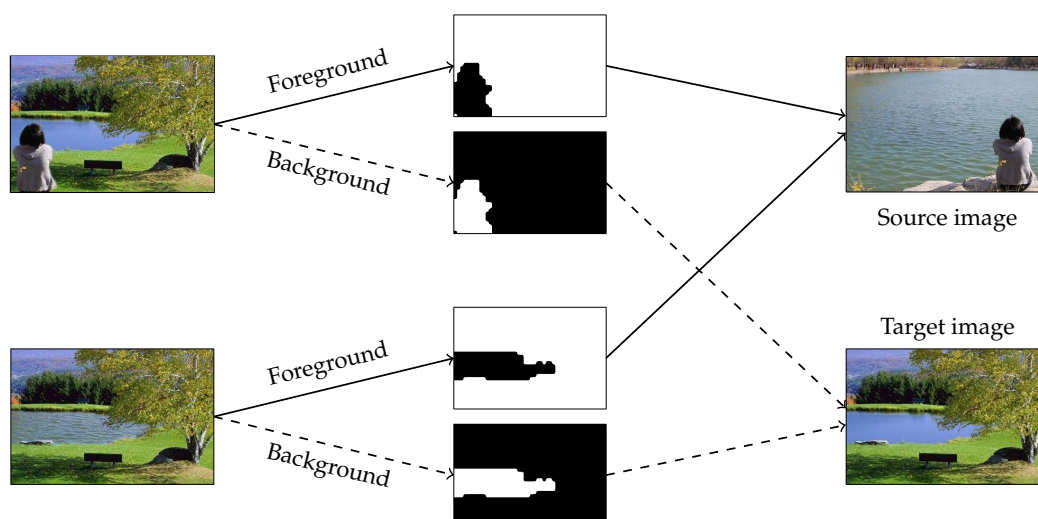


Figure 10. Example of two images verified as spliced images, where the same source image and the same target image were retrieved.

6. Conclusions

This paper solved a problem that seeks original authentic images that were used to compose the spliced images. Prior to the image-retrieval stage, we segmented the spliced images into the spliced regions and background regions in the image-splicing localization stage in order to improve retrieval accuracy. These segmented regions were then separately queried to find the provenance of the spliced image. We proposed a hybrid method that can effectively employ Zernike moment features and SIFT features for image retrieval. Since the former is invariant to rotation and the latter is invariant to scaling and rotation, the proposed method can handle different splicing operations to find the matching features. By retrieving the source and the target images of the spliced images with high accuracy, the proposed method can validate the authenticity of the spliced images with more certainty (see Figure 10). In future works, we will extend our research on improving the performance of image-splicing localization and the retrieval of small and smooth spliced regions.

Author Contributions: Conceptualization, C.-S.P.; data curation, G.-R.K. and C.-S.P.; methodology, N.T.P. and C.-S.P.; project administration, C.-S.P.; resources, J.-W.L.; software, N.T.P. and J.-W.L.; visualization, N.T.P. and G.-R.K.; writing—original draft, N.T.P.; writing—review and editing, J.-W.L., G.-R.K., and C.-S.P.

Acknowledgments: This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2016-0-00312) supervised by the IITP (Institute for Information and Communications Technology Promotion). This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (NRF-2016R1C1B1009682).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. He, Z.; Lu, W.; Sun, W.; Huang, J. Digital image splicing detection based on Markov features in DCT and DWT domain. *Pattern Recognit.* **2012**, *45*, 4292–4299. [[CrossRef](#)]

2. Sheng, H.; Shen, X.; Shi, Z. Image Splicing Detection Based on the Q-Markov Features. In Proceedings of the Pacific-Rim Conference on Multimedia, Hefei, China, 21–22 September 2018; pp. 453–464.
3. Chen, B.; Qi, X.; Sun, X.; Shi, Y.Q. Quaternion pseudo-Zernike moments combining both of RGB information and depth information for color image splicing detection. *J. Vis. Commun. Image Represent.* **2017**, *49*, 283–290. [[CrossRef](#)]
4. Han, J.G.; Park, T.H.; Moon, Y.H.; Eom, I.K. Efficient Markov feature extraction method for image splicing detection using maximization and threshold expansion. *J. Electron. Imaging* **2016**, *25*, 023031. [[CrossRef](#)]
5. Pham, N.T.; Lee, J.W.; Kwon, G.R.; Park, C.S. Efficient image splicing detection algorithm based on Markov features. *Multimed. Tools Appl.* **2018**, *77*. [[CrossRef](#)]
6. Bondi, L.; Lameri, S.; Guera, D.; Bestagini, P.; Delp, E.J.; Tubaro, S. Tampering Detection and Localization through Clustering of Camera-Based CNN Features. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 453–464.
7. Zampoglou, M.; Papadopoulos, S.; Kompatsiaris, Y. Large-scale evaluation of splicing localization algorithms for web images. *Multimed. Tools Appl.* **2017**, *76*, 4801–4834. [[CrossRef](#)]
8. Yao, H.; Wang, S.; Zhang, X.; Qin, C.; Wang, J. Detecting Image Splicing Based on Noise Level Inconsistency. *Multimed. Tools Appl.* **2017**, *76*, 12457–12479. [[CrossRef](#)]
9. Wu, Y.; Abd-Almageed, W.; Natarajan, P. Deep Matching and Validation Network: An End-to-End Solution to Constrained Image Splicing Localization and Detection. In Proceedings of the 2017 ACM International Conference on Multimedia Conference, Mountain View, CA, USA, 23–27 October 2017; pp. 1480–1502.
10. Zheng, L.; Yang, Y.; Tian, Q. SIFT Meets CNN: A Decade Survey of Instance Retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1224–1244. [[CrossRef](#)]
11. Hong, W.; Chen, M.; Chen, T.S. An efficient reversible image authentication method using improved PVO and LSB substitution techniques. *Signal Process. Image Commun.* **2017**, *58*, 111–122. [[CrossRef](#)]
12. Hong, W.; Chen, M.; Chen, T.S.; Huang, C.C. An efficient authentication method for AMBTC compressed images using adaptive pixel pair matching. *Multimed. Tools Appl.* **2018**, *77*, 4677–4695. [[CrossRef](#)]
13. Zeng, H.; Zhan, Y.; Kang, X.; Lin, X. Image splicing localization using PCA-based noise level estimation. *Multimed. Tools Appl.* **2017**, *76*, 4783–4799. [[CrossRef](#)]
14. Chen, C.; McCloskey, S.; Yu, J. Image Splicing Detection via Camera Response Function Analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1876–1885.
15. Wang, P.; Liu, F.; Yang, C.; Luo, X. Blind forensics of image gamma transformation and its application in splicing detection. *J. Vis. Commun. Image Represent.* **2018**, *55*, 80–90. [[CrossRef](#)]
16. Bappy, J.H.; Roy-Chowdhury, A.K.; Bunk, J.; Nataraj, L.; Manjunath, B.S. Exploiting Spatial Structure for Localizing Manipulated Image Regions. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4970–4979.
17. Brogan, J.; Bestagini, P.; Bharati, A.; Pinto, A.; Moreira, D.; Bowyer, K.; Flynn, P.; Rocha, A.; Scheirer, W. Spotting the Difference: Context Retrieval and Analysis for Improved Forgery Detection and Localization. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 4078–4082.
18. Maigrot, C.; Kijak, E.; Claveau, V. Context-aware forgery localization in social-media images: A feature-based approach evaluation. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 545–549.
19. Lin, Z.; He, J.; Tang, X.; Tang, C.K. Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. *Pattern Recognit.* **2009**, *42*, 2492–2501. [[CrossRef](#)]
20. Khotanzad, A.; Hong, Y.H. Invariant Image Recognition by Zernike Moments. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 489–497. [[CrossRef](#)]
21. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
22. Pinto, A.; Moreira, D.; Bharati, A.; Brogan, J.; Bowyer, K.; Flynn, P.; Scheirer, W.; Rocha, A. Provenance filtering for multimedia phylogeny. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1502–1506.

23. Bharati, A.; Moreira, D.; Pinto, A.; Brogan, J.; Bowyer, K.; Flynn, P.; Scheirer, W.; Rocha, A. U-phylogeny: Undirected provenance graph construction in the wild. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1517–1521.
24. Moreira, D.; Bharati, A.; Brogan, J.; Pinto, A.; Parowski, M.; Bowyer, K.W.; Flynn, P.J.; Rocha, A.; Scheirer, W.J. Image Provenance Analysis at Scale. *IEEE Trans. Image Process.* **2018**, *27*, 6109–6123.
25. Khotanzad, A.; Lu, J.H. Classification of invariant image representations using a neural network. *IEEE Trans. Acoust.* **1990**, *38*, 1028–1038. [[CrossRef](#)]
26. Hwang, S.K.; Kim, W.Y. A novel approach to the fast computation of Zernike moments. *Pattern Recognit.* **2006**, *39*, 2065–2076. [[CrossRef](#)]
27. Papakostas, G.A.; Boutalis, Y.S.; Karras, D.A.; Mertzios, B.G. A new class of Zernike moments for computer vision applications. *Inf. Sci.* **2007**, *177*, 2802–2819. [[CrossRef](#)]
28. Papakostas, G.A.; Boutalis, Y.S.; Karras, D.A.; Mertzios, B.G. Pattern classification by using improved wavelet compressed Zernike moments. *Appl. Math. Comput.* **2009**, *212*, 162–176. [[CrossRef](#)]
29. Ryu, S.J.; Lee, M.J.; Lee, H.K. Detection of Copy-Rotate-Move Forgery Using Zernike Moments. In Proceedings of the International Workshop on Information Hiding, Calgary, AB, Canada, 28–30 June 2010; pp. 51–65.
30. Christlein, V.; Riess, C.; Jordan, J.; Riess, C.; Angelopoulou, E. An Evaluation of Popular Copy-Move Forgery Detection Approaches. *IEEE Trans. Inf. Forensic Secur.* **2012**, *7*, 1841–1854. [[CrossRef](#)]
31. Park, C.S.; Kim, C.; Lee, J.; Kwon, G.R. Rotation and scale invariant upsampled log-polar fourier descriptor for copy-move forgery detection. *Multimed. Tools Appl.* **2016**, *75*, 16577–16595. [[CrossRef](#)]
32. Zheng, J.; Liu, Y.; Ren, J.; Zhu, T.; Yan, Y.; Yang, H. Fusion of block and keypoints based approaches for effective copy-move image forgery detection. *Multidimens. Syst. Signal Process.* **2016**, *27*, 989–1005. [[CrossRef](#)]
33. Fahmy, O.M. A new Zernike moments based technique for camera identification and forgery detection. *Signal Image Video Process.* **2017**, *11*, 785–792. [[CrossRef](#)]
34. Liu, Y.; Huang, T. Exposing video inter frame forgery by Zernike opponent chromaticity moments and coarseness analysis. *Multimed. Syst.* **2017**, *23*, 223–238. [[CrossRef](#)]
35. Kim, H.S.; Lee, H.K. Invariant Image Watermark Using Zernike Moments. *IEEE Trans. Circuits Syst. Video Technol.* **2003**, *13*, 766–775.
36. Lutovac, B.; Daković, M.; Stanković, S.; Orović, I. An algorithm for robust image watermarking based on the DCT and Zernike moments. *Multimed. Tools Appl.* **2017**, *76*, 23333–23352. [[CrossRef](#)]
37. Tsougenis, E.D.; Papakostas, G.A.; Koulouriotis, D.E.; Tourassis, V.D. Performance evaluation of moment-based watermarking methods: A review. *J. Syst. Softw.* **2012**, *85*, 1864–1884. [[CrossRef](#)]
38. Kumar, Y.; Aggarwal, A.; Tiwari, S.; Singh, K. An efficient and robust approach for biomedical image retrieval using Zernike moments. *Biomed. Signal Process. Control* **2018**, *39*, 459–473. [[CrossRef](#)]
39. Zhou, W.; Li, H.; Tian, Q. Recent Advance in Content-based Image Retrieval: A Literature Survey. *arXiv* **2017**, arXiv:1706.06064.
40. Jegou, H.; Douze, M.; Schmid, C. Hamming Embedding and Weak Geometric Consistency for Large Scale Image Search. In Proceedings of the European Conference on Computer Vision (ECCV), Marseille, France, 12–18 October 2008; pp. 304–317.
41. Gordo, A.; Almazán, J.; Revaud, J.; Larlus, D. Deep Image Retrieval: Learning Global Representations for Image Search. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 241–257.
42. Gordo, A.; Almazan, J.; Revaud, J.; Larlus, D. End-to-end Learning of Deep Visual Representations for Image Retrieval. *Int. J. Comput. Vis.* **2017**, *124*, 237–254. [[CrossRef](#)]
43. Nguyen, N.B.; Nguyen, K.M.; Van, C.M.; Le, D.D. Graph-based visual instance mining with geometric matching and nearest candidates selection. In Proceedings of the International Conference on Knowledge and Systems Engineering (KSE), Hue, Vietnam, 19–21 October 2017; pp. 263–268.
44. Park, C.S.; Choeh, J.Y. Fast and robust copy-move forgery detection based on scale-space representation. *Multimed. Tools Appl.* **2018**, *77*, 16795–16811. [[CrossRef](#)]
45. Amerini, I.; Ballan, L.; Caldelli, R.; Del Bimbo, A.; Serra, G. A SIFT-Based Forensic Method for Copy-Move Attack Detection and Transformation Recovery. *IEEE Trans. Inf. Forensic Secur.* **2011**, *6*, 1099–1110. [[CrossRef](#)]

46. Zheng, L.; Wang, S.; Liu, Z.; Tian, Q. Packing and Padding: Coupled Multi-index for Accurate Image Retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 1939–1946.
47. Zhou, W.; Li, H.; Hong, R.; Lu, Y.; Tian, Q. BSIFT: Toward Data-Independent Codebook for Large Scale Image Search. *IEEE Trans. Image Process.* **2015**, *24*, 967–979. [[CrossRef](#)] [[PubMed](#)]
48. Sattler, T.; Havlena, M.; Schindler, K.; Pollefeys, M. Large-Scale Location Recognition and the Geometric Burstiness Problem. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1582–1590.
49. Tolias, G.; Avrithis, Y.; Jégou, H. To Aggregate or Not to aggregate: Selective Match Kernels for Image Search. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013; pp. 1401–1408.
50. Arandjelović, R.; Zisserman, A. DisLocation: Scalable Descriptor Distinctiveness for Location Recognition. In Proceedings of the Asian Conference on Computer Vision (ACCV), Singapore, 1–5 November 2014; pp. 188–204.
51. Jégou, H.; Douze, M.; Schmid, C. On the burstiness of visual elements. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 1169–1176.
52. Robertson, S. Understanding inverse document frequency: On theoretical arguments for IDF. *J. Doc.* **2004**, *60*, 503–520. [[CrossRef](#)]
53. Dong, J.; Wang, W.; Tan, T. CASIA Image Tampering Detection Evaluation Database. In Proceedings of the IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP), Beijing, China, 6–10 July 2013; pp. 422–426.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).