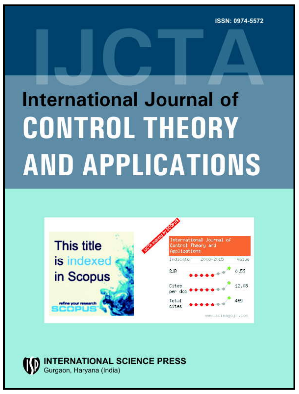


International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 13 • 2017

Copy-Move Forgery Detection using Histogram Based Package Clustering and Sorted Consecutive Local Binary Patterns

V. M. Manikandan[1] and V. Masilamani[2]

[1] *Department of Computer Science and Engineering Indian Institute of Information Technology Design and Manufacturing Kancheepuram, Chennai, Tamilnadu, Email- coe14d001@iiitdm.ac.in*
[2] *V. Masilamani Department of Computer Science and Engineering Indian Institute of Information Technology Design and Manufacturing Kancheepuram, Chennai, Tamilnadu, Email- masila@iiitdm.ac.in*

Abstract: Copy-move forgery is one of the most common image tampering attacks on digital images. In this paper, we present a block-based copy-move forgery detection method based on sorted consecutive local binary pattern(SCLBP) with package clustering based on the distribution of pixels in the local histograms. In the proposed scheme original images will be divided into overlapping blocks, and each block will put into a suitable package by analyzing the histogram. The number of comparisons has been reduced by comparing the feature vectors the blocks within the same package and with the blocks in very adjacent package alone. To reduce the false positive results further refinement on the detection result has been done by removing the smaller connected components. Experimental results show that the proposed scheme outperforms the existing scheme in terms of detection accuracy.

Keywords: copy and move, forgery, sorted consecutive local binary pattern, package clustering

1. INTRODUCTION

Nowadays digital image transmission has been a common practice due to the advancement in communication technologies and cheap availability of multimedia devices with image capturing facility. There are lots of image editing software available to process the digital images by different means such as resizing the image, cropping of unwanted regions of the image, improving the brightness, etc. In general, image editing software such as Adobe Photoshop, Adobe Lightroom, Serix Photo Plus etc. have been used to improve the visual effects of the images. But there is a wide scope to use this software to edit digital images in such a way that the edited images can be used as false evidences.

Over the past few years, many techniques have been introduced to tamper images, and these techniques are mainly classified into the following three categories:

1. *Copy-move forgery* : In this, a part of an image is copied and pasted over another part of the same image to conceal some regions.

2. *Image splicing* : A process which combines two or more images to create a new image.
3. *Image retouching* : Alteration of an image for its enhancement, or for the introduction of desired features, or for the reduction of undesired features.

Digital image forensics is a field that establishes the credibility and authenticity of images by different ways. This field is fast growing due to its wide applicability in different domains such as sports, news reporting, insurance claiming etc. Present digital image forensic approaches are classified into *active* or *passive* techniques. In the case of active methods, a *digital watermark* or a *digital signature* is embedded into the original image, which can be extracted later to verify the authenticity of the image. In passive forgery detection, no prior information is required about the source image, and this method works based on the fact that there will be some statistical change in the digital images that can be used to identify different kinds of forgery[1, 2, 3, 4].

In this work, we considered forgery detection methods to detect copy-move forgery. Generally, Copy-move forgery detection techniques can be classified into two categories: *block-based* approaches [5, 6, 7] and *keypoint* based approaches [8, 9, 10]. In block based methods, the suspicious image will be divided into overlapping blocks of the specified size and a feature vector will be computed for these blocks. Image blocks that have similar feature vectors are considered to be regions with copy and move forgery. In key-point based methods, feature vectors are computed only for the keypoint regions in the image and there is no subdivision into blocks. The feature vectors of the selected key points will be compared to identify the copy and move regions in the forged image. Keypoint based copy-move forgery detection may get failed when there are no enough keypoints in the forged regions. In such cases also block-based approach will work better, but the computational complexity of the block-based approach is high.

Recent works in block-based copy and move forgery detection is concerned about the reduction of time complexity without compromising the detection capability. The easiest way to detect such forgery is the exhaustive search for matching pixels, but this approach has two major drawbacks: firstly this method is computationally expensive and secondly, it fails to detect the forgery in case the copied portion has undergone some modifications such as rotation, scaling, smoothing, etc. The exhaustive approach will take $(M.N)^2$ steps for an image of $M{\times}N$ pixels[11]. Later, different approaches are proposed in this area to reduce the computation time with better copy-move detection capability.

A method based on multiresolution local binary pattern has been proposed in [12]. In this work first, the image is divided into overlapping blocks and for each block, LBP patterns are extracted as the feature vector. Further, the feature vectors are sorted based on the lexicographical order to reduce the computational complexity. Duplicated image blocks are determined in the block matching step using k-d tree for more time reduction. Another discrete cosine transform(DCT) based copy-move detection method has been proposed in [13], in which authors used the low frequency coefficients to reduce the size of feature vectors. Similar to the procedure in [12] the feature vectors are further lexicographically sorted to bring similar features nearby positions to reduce the time complexity of feature matching. In general due to energy compaction property of DCT, DCT based approaches are more robust as compared to the principal component analysis (PCA) based approach proposed in [14]. Later combination of DCT-PCA also has been proposed in [15] to improve the copy-move detection capability.

Recently, discrete cosine transform(DCT) and package clustering based method has been proposed in [16] to reduce the number of feature vector comparisons by putting overlapping blocks into 64 different packages based on the pixel mean value. In this work, the authors assumed that the feature vectors of overlapping blocks with almost same pixel mean value only need to be compared to detect the copied and moved regions. Mean values obtained from each 4×4 DCT coefficients of an 8×8 image block has been used as feature vectors. The results from the existing scheme [16] shows that they are achieving better detection of copy and move in less time compared to the other existing methods. In our point of view, there is a high probability that two blocks with entirely different property may come to same package due the same pixel mean, and all such blocks need to be compared after finding out the corresponding feature vectors.

In the proposed work a novel histogram based package clustering method has been introduced to reduce the number of false positive classifications of blocks into same packages during the creation of packages. The recently introduced feature extraction method known as the *consecutive sorted local binary pattern* (SCLBP) [17] has been used for feature comparison. Experimental study shows that the proposed scheme outperforms the recent block based copy and move forgery detection methods in terms of detection rate and time complexity.

The further sections of this paper are organized as follows : Section II discusses about the proposed scheme and Section III discusses about the experimental study and result analysis. Conclusion and future works is given in Section IV.

2. PROPOSED SCHEME

An overview of the proposed scheme is shown in Figure. 1 and Algorithm 1 describes the sequence of operations in detail.

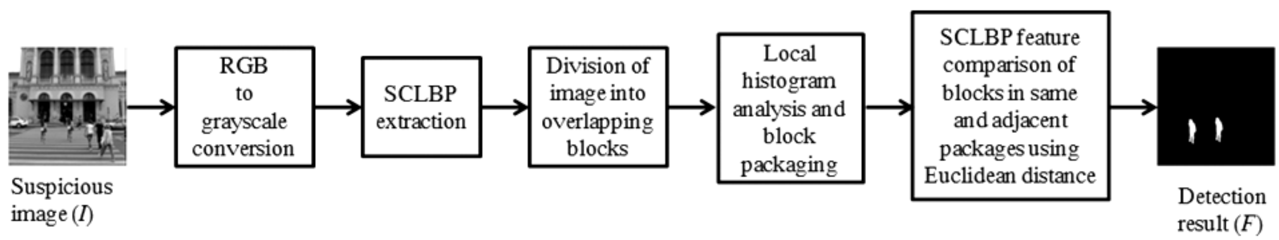


Figure 1: Overview of proposed copy-move forgery detection schme

As per the proposed scheme the given suspicious RGB color image *I* of size $M \times N$ pixels need to be converted into the corresponding 8-bit grayscale image. Processing grayscale image is computationally less complex, as compared to RGB image. In RGB representation each color pixel at spatial location (*x, y*) will be represented by using three different color planes. Let us assume that r_{xy}, g_{xy}, and b_{xy} are the respective R, G, and B color components the corresponding grayscale value G_{xy} can be computed by equation(1).

$$G_{xy} = (0.299 \times r_{x,y}) + (0.587 \times g_{x,y}) + (0.114 \times b_{x,y}) \quad (1)$$

Next step is to find out the SCLBP feature vector for every pixel in the grayscale image *G*' by considering all the 8-neighborhood pixels. To handle the border pixels in the grayscale image *G* we padded 0's around all the four sides of the image, which is considered as *G'*.

The next step is to divide the image into overlapping blocks of size $B \times B$ pixels and put them into suitable packages. If we consider *P* number of packages, and to find the suitable package PA_k for a given block C_i, first find the local histogram H_i of the block C_i. Further, divide the histogram into equal regions of size . An illustrative example is shown in Figure. 2, when we considered *P* is equal to 16.

As shown in Figure 2. the local histogram H_i of block C_i will be divided into 16 regions, and the package number is determined by the region which contain more number of pixels. For example, if the local histogram H_1 of block C_1 contains more pixels from region 3, then it will go to PA_3. All the overlapping blocks will be distributed into the corresponding packages. In general, we can say that the majority of pixels in the blocks of in a given package has almost same intensity value.

For feature matching we considered all the blocks in the same package and the just adjacent package, if the distance between those blocks are more than a threshold spatial distance T_1. In our work T_1 is empirically

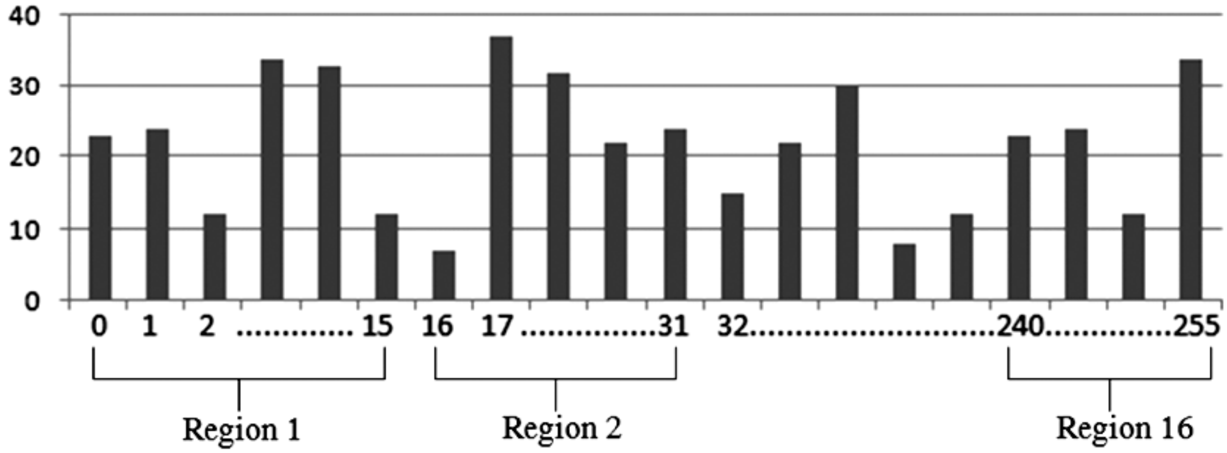


Figure 2: Division of histogram into regions

computed as (N×0.05). This ensures that a large smooth region will not be wrongly classified as copied and moved region. Our assumption is that in general, duplicate regions formed through copy and move forgery will stay in a minimum distance. SCLBP feature vectors of all the pixels in a block will concatenate to generate the feature vector for a specific block. The feature vectors of all the blocks in the same package and adjacent package will be compared using Euclidean distance. If the feature vectors of two different blocks are matching it will be considered as copy-moved region, and in F all the corresponding position will be replaced with 1. Euclidean distance D between two feature vectors V_1 and V_2 can be calculated based on the equation (2).

$$D(V_1, V_2) = \sqrt{\sum_{i=1}^{L} [V_1(i) - V_2(i)]^2} \text{, where } L \text{ is the length of the feature vectors } V_1 \text{ and } V_2 \quad (2)$$

The binary image F was initialized with 0's and which will get updated during the comparison process. If two pixels in (x_1, y_1) and (x_2, y_2) are identified as copied and moved, then $F(x_1, y_1)$ and $F(x_2, y_2)$ will be updated to 1. Once all the comparisons get over, to avoid the false positives we will remove all the connected components smaller than the threshold T_2. The threshold value for T_2 is empirically decided as $\left\lfloor \frac{1}{N^2} \right\rfloor \times 100$.

Algorithm 1 : Proposed algorithm to detect copy and move forgery

---

Input : Suspicious RGB color image I of size $N \times N$ pixels

Output : Binary image F of size $N \times N$ shows the copy-move regions with 1's and remaining with 0's

Step 1 : Convert color image I into the corresponding 8-bit grayscale image G using equation (1).

Step 2 : Add one pixel width border to image G with 0 pixels to compute SCLBP for all the pixels in the image, say G'.

Step 3 Compute the SCLBP feature vector V_m for all the pixels during row-wise scanning of pixels in G', where $1 \leq m \leq N \times N$, V_m is the SCLBP feature of a single pixel.

Step 4 : Initialize a package cluster with P number of empty packages, say $PA_1, PA_2, \ldots, PA_P$.

Step 5 : Divide the image G into overlapping blocks of size $B \times B$ pixels, and there will be $(N-B+1)^2$ number of such blocks. Each block $C\text{-}_i$ will go into any one of the packages PA_k based on the distribution of pixel values in the local histogram H_i. H_i will be divided into equal regions, say $R_1, R_2,..,R_P$ of size $\left\lfloor \frac{256}{P} \right\rfloor$. Assume that $S_1, S_2,.., S_P$ are the number of pixels in the regions $R_1, R_2,..,R_P$, and find m such that S_m is greater than or equal to all S_n, 1*d"nd"P*. In this regard block C_i will go to the m^{th} package. Note that here $1 \leq k \leq P$, and $1 \leq i \leq (N-B+1)^2$

Step 6 : Consider all the pairs of blocks with spatial distance more than T_1 in *G'* in the package PA_i and package PA_{i+1} at a time, and compare the SCLBP feature vector of the blocks using Euclidean distance mentioned in equation(2), where $1 \leq i \leq P-1$. The SCLBP feature vector of a block can be computed by row-wise concatenation of the SCLBP features of the pixels in the block. Here T_1 is the spatial distance threshold, which is empirically obtained as ($N\times 0.05$).

Step 7 : If the similarity between SCLBP feature vectors V_p and V_q of two different blocks are matched, and which are located at spatial locations (x_1, y_1) and (x_2, y_2), then update $F(x,y) = 1$ and $F(p, q) = 1$.

Step 8 : Remove all the connected components in the image F with less than T_2 number of pixels.

Step 9 : Output *F*.

3. EXPERIMENTAL STUDY AND RESULT ANALYSIS

The proposed scheme has been evaluated on the tampered images generated by copy-move forgery using Adobe photoshop. The images considered for copy-move forgery has been downloaded from the McGill calibrated color image database [18]. During experimental study the input images are converted into the grayscale images of size 256×256 pixels. The efficiency of a copy-move forgery detection algorithms are measured in terms of detection accuracy, and detection time. During experimental our proposed scheme is compared with the recently proposed well-known forgery detection method which used DCT as the feature vector with block mean for package clustering[16].

1. *Detection accuracy (DA)* : Detection Accuracy is defined as the percentage of forged pixels in an image, correctly detected by a particular detection algorithm.

$$DA = \frac{CCMP}{ACM} \times 100 \tag{3}$$

Where CCMP denotes the total number of correctly detected copy-moved pixels, ACM denotes the actual number copy-moved pixels.

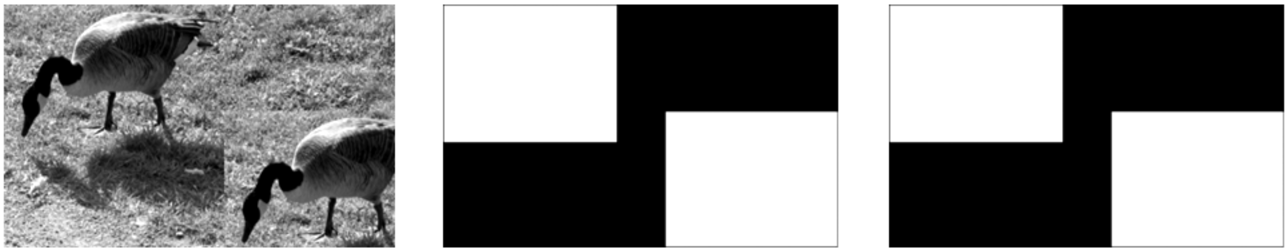
2. *Detection time :* Denotes the time taken by forgery detection algorithm to generate the location map(binary image) to show the copy-moved regions in the given image.

3.1. Detection accuracy

A sample results obtained from both existing scheme and proposed scheme along with detection accuracy is shown Figure. 3. From the existing scheme detection accuracy(DA) is obtained as 92.35%, and from the proposed scheme we achieved DA of 93.15% for the forged image shown in Figure 3(a).

3.2. Selection of block size and package size

The two main parameters that will determine the detection rate and the detection time of the proposed scheme is the block size ($B\times B$), and the package size (P). The total number of overlapping blocks will decrease while we are increasing the block size. If the number of blocks from an image is less then comparison time will be



(a) Forged image **(b) Detection results from proposed scheme(DA : 93.15%)** **(c) Detection results from Existng scheme [16](DA : 92.35%)**

Figure 3: Comparison of copy-move detection accuracy

decreased. If the package size is too small then the region size during histogram analysis will be wide and blocks with a large pixel difference also will come into the same package. We empirically decided that a block size of size 8×8 pixels with package size of 64 is giving a good trade-off between detection time and detection rate.

The block classification into different packages will help to reduce the number of feature vector comparisons. The optimum feature vector comparison will happen when the blocks are equally distributed in all the packages. The distribution of blocks into packages are completely dependent on two parameters: The pixel distribution in the image, and the way of packaging. In this work we proposed a new packaging scheme for copy-move detection by analyzing the histogram of every block, and the distribution of blocks into packages is analyzed here.

Let $f_k(i)$ is the number of blocks in the package i for the image k, where $1 \leq i \leq P$.

Where M is the number images of taken for experimental study (4)

Claim: f(i) is almost flat.

The flatness of a function f is defined as the standard deviation of $\{f(i) \mid i=1, 2, ..., P\}$, where P is the number of packages. For instance, if $f(i)=a$, then the standard deviation of $\{f(i) \mid i=1, 2, ..., P\}$ is equal to 0.

The flatness property of block distribution in packages has been measured for 420 images randomly selected from [18]. For the proposed scheme we achieved a flatness measure of 347.21, where the flatness measure obtained from existing scheme [16] is 474.86. If the flatness measure is less means the number block comparison will happen during copy-move detection is less, and the flatness measure for proposed scheme is less as compared to existing scheme.

4.3. Time complexity analysis of proposed scheme

Let us assume that original RGB color image I contain $N \times N$ pixels. In the proposed scheme the major time taking operations are RGB to grayscale conversion, SCLBP feature extraction, package generation, and feature matching. For RGB to grayscale conversion, every pixel in I need to be accessed at most once, hence the operation will take $O(N^2)$ time. Further, for computation of SCLBP again all the pixels in the grayscale version of I, say G' need to be accessed only once. In this regard, we can say feature vector generation can be carried out in $O(N^2)$ time. Further package generation operation should consider all the possible overlapping blocks in the image, there will be $(N\text{-}B+1)^2$ number of such blocks, and the local histogram analysis for every block can be carried in $O(N^2)$ time, hence the block size is fixed to B which is independent of the input size and the total number blocks in the image is $(N\text{-}B+1)^2$. Once the packaging of blocks has been finished, the next step is to compare the features of pixels in the same and adjacent packages. The comparison will be efficient when the number of blocks is equally distributed to all the P packages. The worst case performance will happen when all the $(N\text{-}B+1)^2$ number of blocks may go into the same package. In such situation we need to compare feature vectors of S number of block pairs. Where S can be computed by the equation (5).

$$S=qC2, \text{ where } q=(N-B+1)^2 \tag{5}$$

In worst case there will be *qC2* pairs of blocks need to be compared and the feature vector comparison can be carried out with a time complexity of $O(N^4)$. In this regard the time complexity of the proposed copy-move forgery detection method is computed as $O(N^4)$.

Comparison of theoretical time complexity for copy-move detection of the proposed scheme with existing scheme are shown in Table 1.

Table 1
Comparison of time complexity

	Theoretical time complexity
Existing scheme[16]	$O(N^4)$
Proposed scheme	$O(N^4)$

4. CONCLUSION AND FUTURE WORKS

A block-based copy-move forgery detection algorithm is proposed in this paper. To reduce the number of block comparison a new way of block packaging has been introduced in this paper by analyzing the histogram. A recently proposed feature vector known as sorted consecutive local binary pattern(SCLBP) has been used for the comparison of blocks. Experimental studies on the tampered image shows that proposed scheme having a good detection rate. To achieve the complete benefits package clustering in copy-move forgery detection, the blocks should be equally distributed in all the packages. The fixed packaging concept that we used in the proposed scheme will not ensure the same, and hence an improvement can be done over this method. The package number and packaging criteria should be dynamically determined based on the property of the image to achieve almost equal distribution of blocks into packages. The proposed scheme still not be able to provide real-time performance, which can be considered in future work.

REFERENCES

[1] Wang, Xiang-Yang, et al. "A new keypoint-based copy-move forgery detection for small smooth regions."*Multimedia Tools and Applications*(2016): 1-30.

[2] Li, Jingwei, et al. "Keypoint-based copy-move detection scheme by adopting MSCRs and improved feature matching." *Multimedia Tools and Applications*(2016): 1-15.

[3] Sachdev, Kanica, Mandeep Kaur, and Savita Gupta. "A Robust and Fast Technique to Detect Copy Move Forgery in Digital Images Using SLIC Segmentation and SURF Keypoints." *Proceeding of International Conference on Intelligent Communication, Control and Devices.* Springer Singapore, 2017.

[4] Lee, Jen-Chun, Chien-Ping Chang, and Wei-Kuei Chen. "Detection of copy–move image forgery using histogram of orientated gradients." *Information Sciences* 321 (2015): 250-262.

[5] Lee, Jen-Chun. "Copy-move image forgery detection based on Gabor magnitude."*Journal of Visual Communication and Image Representation* 31 (2015): 320-334..

[6] Hashmi, Mohammad Farukh, and Avinash G. Keskar. "Block and Fuzzy Techniques Based Forensic Tool for Detection and Classification of Image Forgery." *Journal of Electrical Engineering and Technology* 10.4 (2015): 1886-1898.

[7] Moradi-Gharghani, Hajar, and Mehdi Nasri. "A new block-based copy-move forgery detection method in digital images." *International Conference on Communication and Signal Processing (ICCSP*, 2016:1208-1212.

[8] Wang, Xiang-Yang, et al. "A new keypoint-based copy-move forgery detection for small smooth regions."*Multimedia Tools and Applications*(2016): 1-30.

[9] Li, Jingwei, et al. “Keypoint-based copy-move detection scheme by adopting MSCRs and improved feature matching.” *Multimedia Tools and Applications*(2016): 1-15.

[10] Sachdev, Kanica, Mandeep Kaur, and Savita Gupta. “A Robust and Fast Technique to Detect Copy Move Forgery in Digital Images Using SLIC Segmentation and SURF Keypoints.” *Proceeding of International Conference on Intelligent Communication, Control and Devices.* Springer Singapore, 2017, 787-793.

[11] Nathalie Diane, Wandji Nanda, Sun Xingming, and Fah Kue Moise. “A Survey of Partition-Based Techniques for Copy-Move Forgery Detection.”*The Scientific World Journal* 2014 (2014).

[12] Davarzani, Reza, et al. “Copy-move forgery detection using multiresolution local binary patterns.” *Forensic science international* 231.1 (2013): 61-72.

[13] Kumar, Sunil, Jagannath Desai, and Shaktidev Mukherjee. “A fast DCT based method for copy move forgery detection.”*Image Information Processing (ICIIP), 2013 IEEE Second International Conference on.* IEEE, 2013.

[14] A.Popescu, H. Farid, “Exposing digital forgeries by detecting duplicated image regions”, Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2004-515, 2004.

[15] Sunil, Kumar, Desai Jagan, and Mukherjee Shaktidev. “DCT-PCA based method for copy-move forgery detection.”*ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol II.* Springer International Publishing, 2014.

[16] Wang, Huan, et al. “A passive authentication scheme for copy-move forgery based on package clustering algorithm.” *Multimedia Tools and Applications*(2016): 1-18.

[17] Ryu, Jongbin, Sungeun Hong, and Hyun S. Yang. “Sorted consecutive local binary pattern for texture classification.”*IEEE Transactions on Image Processing* 24.7 (2015): 2254-2265.

[18] McGill Calibrated color image database, “http://tabby.vision.mcgill.ca/html/welcome.html” (accessed 25/11/2016).