A DCT-based multiscale binary descriptor robust to complex brightness changes

Sinem Aslan
International Computer Institute
Ege University
Izmir, Turkey
sinem.aslan@ege.edu.tr

Mehmet Yamaç
Electrical & Electronics Engineering
Boğaziçi University
Istanbul, Turkey
mehmet.yamac@boun.edu.tr

Bülent Sankur
Electrical & Electronics Engineering
Boğaziçi University
Istanbul, Turkey
bulent.sankur@boun.edu.tr

***Abstract*—Binary descriptors have been very popular in recent years. One reason is that the algorithms that use them become computationally and memory-wise efficient. Furthermore, they tend to have some inherent robustness against some geometrical variations and against various brightness changes. These changes might result from both internal factors and external factors such as location of the light source, viewing angle, scene properties. In this paper, we describe a binary descriptor which proves to be robust to complex brightness changes such as gamma correction, noise and photometric distortions. The experimental results demonstrate that performance of the descriptor in object recognition and local image analysis tasks.**

Keywords - binary descriptor; robustness to photometric distortion and brightness change;

I. INTRODUCTION

Mobile devices, such as smartphones and tablets, are increasingly used to run image understanding tasks in various applications. Examples include but not limited to object recognition for guidance in museums [1] and in-store shopping [2], matching for outdoors augmented reality [3], and detection of urban objects [4]. Image features and descriptors for these mobile applications must satisfy the constraints of limited memory and processor capacity. More importantly, since such imaging applications are typically under uncontrolled and real-life conditions, descriptors need to be robust to challenging illumination conditions, to distortions such as defocus and motion blur as well as geometric transformations.

Recently, binary descriptors have attracted some attention, not only due to their computational simplicity and memory-efficiency, but also, in some cases, due to their inherent robustness against image variability. One interesting class of binary descriptors result from a sequence of intensity-level comparisons within pixel patches, where greater-than and smaller-than types of observations are converted to logical 1 and 0's. These methods essentially probe the gray-level slope configuration around the patch center. The popular ones in the current literature are BRIEF [5], ORB [6], and BRISK [7]. These mainly differ from each other in (*i*) the geometrical pattern with which the pixel pairs are tested, e.g., whether they follow a pseudo-random pattern [5,6] or a specific crafted pattern [7], (ii) the choice of pixel pairs upon which comparison tests are made, i.e. if the pairs to be selected were learned [6] or not [5,7], (iii) and if the orientation compensation as a preprocessing step was included [6,7] or not [5]. It is reported in [8] that the binary descriptors, ORB and BRISK perform quite well under viewpoint changes, zoom and rotation effects and outperform BRIEF. On the other hand, under brightness changes, blur and jpeg compression, BRIEF outperforms its two competitors, ORB and BRISK. Notice that the sensitivity of all these descriptors to noise has to be mitigated by smoothing the input images.

Among the more recent binary descriptors, one can mention ALOHA [9] and Bi-DCT [10]. ALOHA uses a 3-level comparison of pixels of the local patch and slightly outperforms BRIEF for the same sized feature vector. Bi-DCT [10], originally proposed for dense stereo matching, is the method most akin to our proposed method. We work also on 2D-DCT coefficients as in [10]. However while [10] groups all DCT coefficients in one frequency band, we select in each block size a fixed number of DCT coefficients according to an energy criterion. The binarization scheme in [10] is different in that they consider a two-bit, four-level quantizer, where the quantizer dead-zone corresponds to small amplitudes while the large coefficients have signed quantization. The small perturbations are thus eliminated with a threshold computed based on a Cauchy distribution model at each particular frequency layer. We follow a scheme which generates 1-bit codes by quantizing absolute value of DCT coefficients based on the mean value. This approach favors the selection of large and sparse coefficients**.**

In this paper, we propose a new binary descriptor that is memory-efficient and highly robust to illumination changes and photometric distortions. This simple method can potentially take advantage of hardware for DCT compression and may even be applicable to compressed images directly. We dub it MB-DCT (Multiscale Binary-DCT). We evaluate its performance on databases to demonstrate its robustness against various geometrical and photometrical transformations both in the context of object recognition and interest point matching. Its performance is compared with its nearest competitors, BRIEF and Bi-DCT. We demonstrate that MB-DCT performs quite well in the presence of linear and nonlinear brightness changes and photometric distortions such as blur, noise and compression artefacts.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

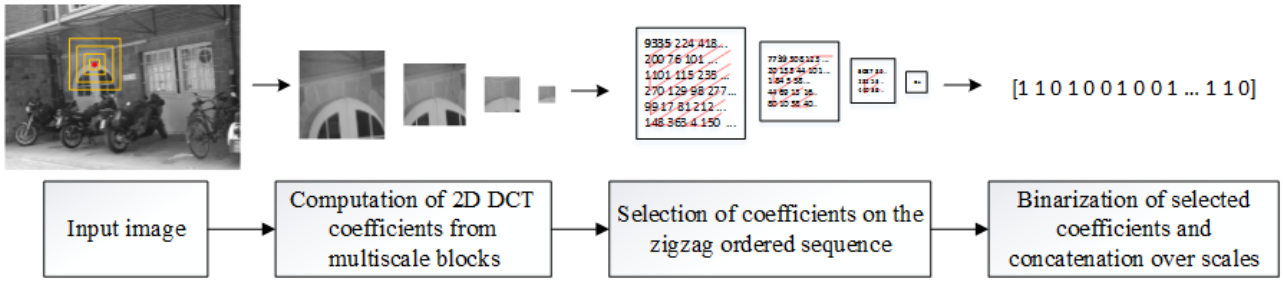

Figure 1. Framework for proposed MB-DCT

We describe the proposed method, MB-DCT, in Section 2. In Section 3, we present the experimental setup used in the evaluation study. We give the experimental results in Section 4, and draw conclusions in Section 5.

II. Proposed Method

Computation of MB-DCT for keypoints that were detected sparsely by a feature detector (i.e. SURF) or densely on a regular grid on images is implemented in three steps as visualized in Figure 1: (i) Feature extraction, (ii) Selection of DCT coefficients, (iii) Binarization of coefficients and their concatenation over scales. Details are given in the following subsections.

A. Feature extraction

DCT is known to have good energy compaction property for certain signal classes and a fast transform implementation [11][12]. These advantages of DCT have motivated us to use it in a new robust binary descriptor design. For each keypoint in an image, we delineate R number of concentric blocks, $P_i \in \mathbb{R}^{N_i \times N_i}$ of increasing size $N_i \times N_i$, $i=1, \ldots, R$, each centered on the keypoint, and compute their 2D DCT:

$$F_i(u,v) = \left| \frac{2}{N_i} c(u)c(v) \sum_{x=0}^{N_i-1} \sum_{y=0}^{N_i-1} P_i(x,y) \cos\left(\frac{\pi(2x+1)u}{2N_i}\right) \cos\left(\frac{\pi(2y+1)v}{2N_i}\right) \right| \quad (1)$$

where $c(k) = 1/\sqrt{2}$ if k = 0, and 1 otherwise; $u,v = 0, 1, \ldots, N_i-1$ and N_i^2 constitutes to the number of pixels in the block of scale i. In the R sets of $N_i \times N_i$ DCT coefficients $\{F_1(u,v)\}_{u,v=0}^{N_1-1}, \{F_2(u,v)\}_{u,v=0}^{N_2-1}, \ldots, \{F_R(u,v)\}_{u,v=0}^{N_R-1}$ we consider their absolute value, as in Eq. 1.

B. Selection of DCT Coefficients

We want to discard irrelevant DCT coefficients while keeping the more informative ones, in order to provide robustness to photometrical distortions and at the same time to have adequate discriminative capacity. The DC term, $\{F_i(0,0)\}_{i=1}^{R}$ is discarded in all scales to desensitize the feature vectors to illumination level. In the zig-zag ordered $\{F_i(u,v)\}_{u,v=1}^{N_i-1}$ coefficients of a block at scale i, we take the first Γ_i ($\Gamma_i < N_i \times N_i$) number of coefficients and discard the remaining ones. In this study we decide to assign Γ_i to scales $i = 1$ to $i = R$ as follows: For the first block of size $N_1 \times N_1$, we extract the first Γ_1 coefficients that correspond to a certain energy percentage. This energy percentage is chosen experimentally using a subset of images from databases used in Section IV. We iterate on the computations of Γ_i till all blocks are represented more or less with the same energy level and at the same time we satisfy the constraint $\sum_{i=1}^{R} \Gamma_i = L$, where L is the length of the descriptor. To compare with other methods such as BRIEF where the feature size is 256, L is selected as 256. Thus, letting Z_i^j denote the j^{th} DCT coefficient in the zig-zag ordered list at scale i, the coefficient vector for that particular scale becomes $\boldsymbol{f}_i = [Z_i^1, Z_i^2, \ldots Z_i^{\Gamma_i}]$.

C. Construction of a Binary Decriptor

Finally, we binarize the selected and ordered $\boldsymbol{f}_i$ DCT vectors by *mean quantization*. Assume that $\{\mu_1, \ldots, \mu_R\}$ are the mean values of the selected DCT coefficients at scales 1 to R, then the final binary descriptor computed for each scale i will be as in Eq. 2 and where $j = 1, 2, \ldots, \Gamma_i$.

$$b_i^j = \begin{cases} 0 & \text{if } Z_i^j < \mu_i \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

Here, mean quantization is used since it is sensitive to large coefficients, so that sparser but possibly more discriminative coefficients will pass the thresholding test. The final binary descriptor for a given keypoint is obtained by concatenation of binarized DCT coefficient sets at each scale $[b_1^{1:\Gamma_1}\ b_2^{1:\Gamma_2} \ldots\ b_R^{1:\Gamma_R}]$.

III. Experimental Setup

Oxford dataset. We first evaluated the proposed descriptor on the Oxford dataset [13] to demonstrate robustness of MB-DCT under the transformations of blur, illumination changes, viewpoints changes, and jpeg compression. Oxford dataset has been a standard dataset to evaluate descriptors' capabilities under geometric and

Figure 2. Synthesized images from image 6 of "Leuven" class; (a) squared brightness change, (b) square rooted brightness changes.

photometric transformations. We used six image sets from Oxford, that are corrupted by blur (Bikes, Trees), illumination changes (Leuven), Jpeg compression artefacts (UBC), and viewpoint changes (Wall, Graffiti). Each image set consists of six images with increased degree of mentioned transformations. Additionally, in order to examine the method for more complex nonlinear brightness changes we also created synthetic images of the class 'Leuven' similarly as in [14], that is we min-max normalize the 2nd to 6th images into [0, 1] range, and apply *square* and *square root* operations on them. As an example of synthetic images, the 6th image is presented in Fig. 2.

The built-in SURF implementation in OpenCV is used for keypoint detection. We use the same keypoints for all binary descriptor methods inside the image region that excludes border bands of 64-pixel width. This width corresponded to half of the largest block size (N_R = 128) used in the computation of MB-DCT. We opted for cropping the pass-partout, rather than padding the images, e.g., by symmetric reflection for this size. The number of detected keypoints ranged from 600 to 4000 depending on the test images, which was sufficient to make a reliable evaluation. We used the same Nearest Neighbour evaluation metric as in [5], where we detect *N* keypoints in the first image and infer *N* corresponding points on the second image by using the published ground truth data. We then compute the set of left-right matches of the *2N* descriptors by considering nearest neighbours of one side to the other. If the matched points are within a tolerance of 2 pixels from true locations, we call them as "correct matches". We finally compute the recognition rate as the number of correct matches/all matches. In [5], it is stated that this procedure might artificially increase the recognition rates, however since the same procedure is applied for all kind of descriptors, relative rates are still reliable.

Coil-20 dataset. We evaluate MB-DCT performance for object recognition by implementing a Bag-of-Words type encoding on COIL-20 "processed" corpus [15], which contains 20 object categories, each having 72 images with a 5 degree pose interval between. The images are sized 128×128 pixels, and symmetric padding of 64 pixels is applied to accommodate the pixels on boundaries. We work on dense points in this experiment with *stride* equal to 3 pixels. The testing setup named as *coil20_24* in [16] is followed, that is, the images of each object category with

TABLE I. PHOTOMETRIC DISTORTIONS TO GENERATE TEST IMAGES

(b) Blur: Apply Gaussian filter with $\sigma = 3$.	(c) Additive White Gaussian Noise (AWGN): with σ=110
(d) Contrast decrease: Linearly map intensity values in [0,255] to [88,168].	(e) Contrast increase: Linearly map intensity values in [88,168] to [0,255].
(f) (Linear) Brightness decrease: Subtract r percent of image mean intensity from each pixel, r: 80%.	(g) (Linear) Brightness increase: Add r percent of image mean intensity to each pixel, r: 80%.
(h) (Nonlinear) squared brightness change: Take square of intensities.	(i) (Nonlinear) square rooted brightness change: Take square root of intensities.
(j) JPEG compression: apply with quality parameter 2.	

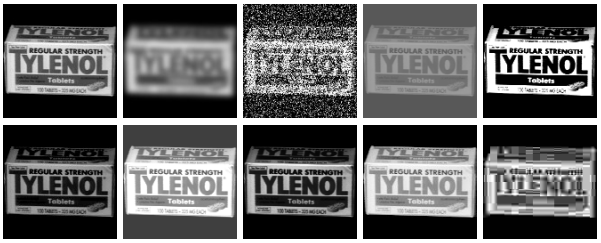

Figure 3. Synthetically applied distortions on a COIL-20 image (*Please see Table 1 for further information*). (a) Original image, (b) Blurring: PSNR=17.7 SSIM=0.5 (c) AWGN: PSNR=10.2 SSIM=0.2 (d) Contrast decr.: PSNR=11.9,SSIM=0.4 (e) Contrast incr.: PSNR=15.1,SSIM=0.9 (f) (Lin.) Brigh. Decr.: PSNR=14.7 SSIM=0.8 (g) (Lin.) Bright. Incr.: PSNR=12.5 SSIM=0.5 (h) (Nonlin.) Sq. Bright.: PSNR=16.4,SSIM=0.9, (i) (Nonlin.) Sq. root. Bright.: PSNR=16.6 SSIM=0.9 (j)JPEG compr.: PSNR=20.5, SSIM=0.8

pose interval of 15 degrees are taken into the training set and the remaining ones into the testing set. We randomly sample a subset from the training set, densely extract MB-DCT vectors from these images and then input them into a clustering algorithm to compute the visual dictionary. For all methods, i.e., BRIEF, Bi-DCT, and MB-DCT, we have used the same chosen subset of training images for building their respective visual dictionary. K-means clustering was used based on Hamming distance and the dictionary size was set at 512. To test the performance of the proposed method in the presence of photometric distortions, we use original images in the training set and apply distortions, as in Table 1, to create test images. A sample image and images under these distortions are presented in Figure 3.

IV. EXPERIMENTS

We evaluated MB-DCT performance for two sizes of the dictionary words, namely 256 and 192 bits, dubbed respectively, MB-DCT-256 and MB-DCT-192. Notice that in both cases, the dictionary size is kept constant at 512 atoms, and the number of concentric blocks is set at R = 6. In particular, the number of DCT coefficients chosen from the R concentric blocks in increasing size were fixed as: $\Gamma_{\text{MBDCT-256}}$={6,16,32,48,64,90} and $\Gamma_{\text{MBDCT-192}}$={16,24,32,35,39,46} number of coefficients are kept after zig-zag ordering of coefficients computed in blocks sized as $N_{\text{MBDCT-256}}$={4,8,16,32,64,128} and $N_{\text{MBDCT-192}}$={8,16,24,32,48,64} respectively for MB-DCT-256 and MB-DCT-192. We implemented Bi-DCT in MATLAB by the default parameters given in [10] and we executed BRIEF-256 (bits) in OPENCV library with the default parameters to make a comparative study.

A. Performance on the Oxford Dataset

The recognition rates for the test sequences Bikes (blur), Trees (blur), Leuven (illumination changes), UBC (jpeg compression), Wall and Graffiti (viewpoint changes) are given in Figure 4. We see that while Bi-DCT-102 is worst among all, proposed MB-DCT-256 and MB-DCT-192 performs quite well in increasing blur distortion even with descriptor length smaller than that of BRIEF-256. While for the brightness changes and synthesized Leuven sequences,

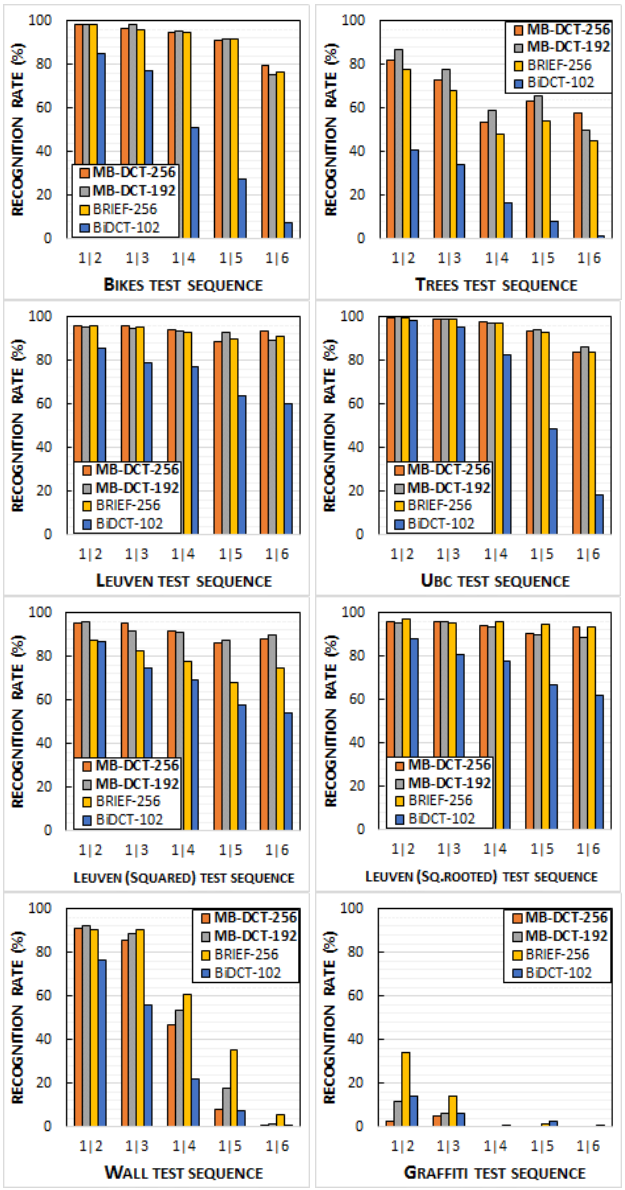

Figure 4. Recognition rates on (a) Bikes (blur), (b) Trees (blur), (c) Leuven (illumination changes), (d) UBC (jpeg compression), (e) Leuven (Squared brightness), (f) Leuven (Sq. rooted Brightness), (g) Wall (viewpoint changes), and (h) Graffiti (viewpoint changes). I|J denotes the recognition rate between image I and image J. Distortion gets monotonically increased from image 1 to image 6.

MB-DCT gives comparative results with BRIEF-256, BRIEF-256 outperforms MB-DCT for viewpoint changes.

B. Object recognition performance on the Coil-20 Dataset

In this experiment we explore the behaviour of the methods for object recognition task when test images were subjected to significant amount of distortion. It should be noticed that filtering is not applied before computing the proposed MB-DCT descriptor and Bi-DCT [10], while box-filtering is applied for BRIEF as default in OpenCV (9×9 sized box filtering for 48×48 sized image patches). Object recognition is accomplished by a simple K-Nearest Neighbour classifier with 5-fold cross validation using chi-square distance. The results are presented in Table 2.

TABLE II. OBJECT RECOGNITION ACCURACY ON COIL20 WHEN PHOTOMETRIC DISTORTIONS ARE APPLIED ON TEST IMAGES.

Modification Type	MBDCT-256	MBDCT-192	BRIEF-256 [5]	BiDCT-102 [10]
No Modification	99.90	99.79	**100**	99.79
Blurring	**94.48**	85.21	75.83	61.67
AWGN	**94.69**	70.21	9.17	5
Contrast Decrease	**99.90**	99.69	**99.90**	99.79
Contrast Increase	**96.35**	94.17	73.96	73.13
(Linear) Bright. decrease	99.79	99.79	**99.90**	85.15
(Linear) Bright. increase	**99.79**	92.50	99.69	81.77
(Nonlin.) Square. Bright.	99.79	99.58	**100**	99.69
(Nonlin.) Sq.root. Bright.	99.79	99.69	**100**	99.79
JPEG compression	**99.48**	98.85	77.19	55.94

For the brightness changes MB-DCT gives comparative results with BRIEF-256. Moreover, MB-DCT performs quite well when distortions such as noise, blur and jpeg compression exist without applying filtering even in lower lengths, e.g. MB-DCT-192.

V. CONCLUSION

In this study, we have proposed a binary descriptor called MB-DCT and evaluated its performance on Oxford dataset and COIL20 object recognition datasets. We demonstrated that the proposed method is highly robust to blur and noise artefacts and gives results comparable to BRIEF in the presence of complex brightness changes even for shorter descriptor lengths. MB-DCT method, however, is not designed to be robust to geometrical transformations such as rotation and viewpoint changes. We are presently investigating a max-pooling approach of DCT coefficients from rotated patches for rotational invariance.

REFERENCES

[1] P. Föckler, T. Zeidler, B. Brombach, E. Bruns, and O. Bimber, "PhoneGuide: museum guidance supported by on-device object recognition on mobile phones", in Proc. of the 4th int. ACM conf. on Mobile and ubiquitous multimedia, pp. 3-10, 2005.

[2] Y. Xu, M. Spasojevic, J. Gao, and M. Jacob, "Designing a vision-based mobile interface for in-store shopping", in Proc. of the 5th ACM Nordic conf. on Human-computer interaction: building bridges, pp. 393-402, 2008.

[3] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W. C. Chen, T. Bismpigiannis, … and B. Girod, "Outdoors augmented reality on mobile phone using loxel-based visual feature organization", in Proc. of the 1st ACM int. conf. on Multimedia information retrieval (pp. 427-434). 2008.

[4] G. Fritz, C. Seifert, and L. Paletta, "A mobile vision system for urban detection with informative local descriptors", IEEE conf. on Comp. Vision Systems (ICVS'06), 2006.

[5] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features", Proc. of European Conference on Computer Vision (ECCV 2010), pp. 778-792, 2010.

[6] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF", Proc. of Int. Conf. on Computer Vision (ICCV 2011), pp. 2564-2571, 2011.

[7] S. Leutenegger, M. Chli, and R. Y. Siegwart, “BRISK: Binary robust invariant scalable keypoints”, Proc. of Int. Conf. on Computer Vision (ICCV 2011), pp. 2548-2555, 2011.

[8] J. Heinly, E. Dunn, and J. M. Frahm. "Comparative evaluation of binary features." Proc. Of European Conference on Computer Vision, pp. 759-773. Springer Berlin Heidelberg, 2012.

[9] S. Saha and V. Demoulin, “ALOHA: An efficient binary descriptor based on Haar features”, Proc. of 19th IEEE Int. Conf. on Image Processing (ICIP 2012), pp. 2345-2348, 2012.

[10] M. J. Sheu, P. Y. Lin, J. Y. Chen, C. C. Lee, and B. S. Lin, “Bi-DCT: DCT-based Local Binary Descriptor for Dense Stereo Matching”, IEEE Signal Processing Letters, 22(7): 847-851, 2015.

[11] G. K. Wallace, “The JPEG still picture compression standard”, IEEE Trans. on Consumer Electronics, 38(1): xviii-xxxiv,1992.

[12] A. C. Hung and TH-Y Meng, “A Comparison of fast DCT algorithms,” Multimedia Systems, 5(2), 1994.

[13] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors”, Trans. Pattern Anal. Mach. Intell., 27(10):1615–1630, 2005.

[14] F. Tang, S. H. Lim, N. L. Chang, and H. Tao, “A novel feature descriptor invariant to complex brightness changes”, Proc. of Comp. Vis. and Pattern Recog. (CVPR 2009), pp. 2631-2638, 2009.

[15] S. A. Nene, S. K. Nayar and H. Murase, "Columbia Object Image Library (COIL-20)," Technical Report CUCS-005-96, February 1996.

[16] S. Aslan, C.B. Akgül, B. Sankur, and E.T. Tunalı, “SymPaD: Symbolic Patch Descriptor”, in Proc. of 10th Int. Conf. on Comp. Vis. Th. and Appl. (VISAPP 2015), pp. 266-271, 2015.

[17] X. Qi, R. Xiao, C. G. Li, Y. Qiao, J. Guo and X. Tang, "Pairwise Rotation Invariant Co-Occurrence Local Binary Pattern," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 11, pp. 2199-2213, Nov. 1 2014.