A DCT-based multiscaled binary descriptor invariant 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 due to their computation and memory efficient capabilities. Many of them could deal with geometrical variations and linear intensity shifts, however, they were not examined for more complex brightness changes specifically, which might result of some internal conditions such as nonlinear camera capturing parameters or external conditions such as location of the light source, viewing angle, scene properties. Moreover, the existing binary descriptors are sensitive to photometric distortions such as noise and blur. In this study, our aim is developing a binary descriptor which is highly robust to complex brightness changes and photometric distortions. The preliminary results demonstrate the noticeable performance of the proposed method when the mentioned distortions were artificially added on testing images of an object recognition dataset.**

Keywords-binary descriptor; image understanding; memory-efficient; robust to photometric distortions

I. Introduction

Mobile devices, e.g., 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], detection of urban objects [4]. Due to memory and processor constraints of such devices, the first requirement is to extract image features that take small memory space. Moreover, since the media capturing technology in these devices is dedicated for real-life image capturing conditions by the average user, features are needed that are robust to challenging illumination conditions or to distortions such as focus and motion blur in order to achieve image understanding tasks successfully.

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 shape slope configuration around the patch center. The popular ones in the current literature are BRIEF [5], ORB [6], BRISK [7], FREAK [8]. 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* [7,8], (*ii*) the choice of pixel pairs upon which comparison tests are made, i.e. *if the pairs to be selected were learned* [6,8] *or not* [5,7], (*iii*) and if the *orientation compensation as a preprocessing step was included* [6,7,8] or *not* [5]. It is reported in [9] that these three binary descriptors, FREAK, ORB and BRISK perform quite similarly under viewpoint changes, zoom and rotation effects. However, under brightness changes, blur and jpeg compression, BRIEF outperforms its two competitors. Notice that the sensitivity of these descriptors to noise has to be mitigated by smoothing the input images.

Among more recent binary descriptors one can mention ALOHA [10] and Bi-DCT [11]. ALOHA uses a 3-level comparison of pixels of the local patch and it slightly outperforms BRIEF for the same sized feature vector. Bi-DCT [11], originally proposed for dense stereo matching, is the method most akin to our proposed method. We similarly work on 2D DCT coefficients similar to [11], however [11] resides in the same scale and their binarization scheme is different than ours. In [11], a test with three comparison steps which generates two-bit grey-codes for each DCT coefficients is performed. The perturbations are eliminated with respect to a threshold computed by a function based on the Cauchy distribution of the particular frequency layer. We follow a simpler scheme which generates 1-bit grey-codes, comparisons are made with respect to the mean of DCT coefficients after eliminating the ones related to perturbations.

In this paper, we propose a new binary descriptor that is memory-efficient and highly robust to illuminations changes and photometric distortions. The method is simple and can potentially take advantage of hardware for DCT compression and may even be applicable to compressed images directly. We name it as MB-DCT (Multiscale Binary-DCT). We evaluate its performance on Oxford dataset to demonstrate its robustness against various geometrical and photometrical transformations and on COIL-20 dataset for object recognition task. Its performance is compared with its nearest competitors, BRIEF and Bi-DCT. We demonstrate that MB-DCT performs quite well in the presence of complex brightness changes and photometric distortions such as blur, noise and compression artefacts.

We describe the proposed method, MB-DCT, in Section 2. In Section 3, we present the experimental setup used in the

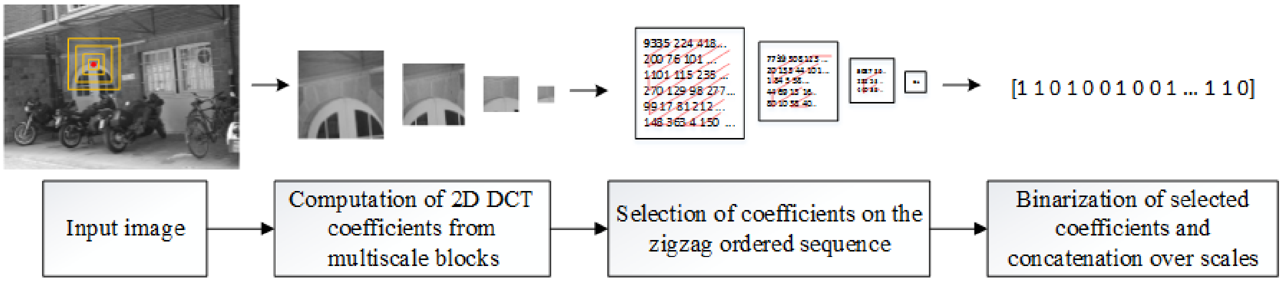
[1 1 0 1 0 0 1 0 0 1 ... 1 1 0]
Input image
Computation of 2D DCT coefficients from multiscale blocks
Selection of coefficients on the zigzag ordered sequence
Binarization of selected coefficients and concatenation over scales

Figure 1. Framework for proposed MB-DCT

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) elimination of useless 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 [12][13]. These advantages of DCT have motivated us to use it in a new robust binary descriptor design. For each keypoint in an image, we designate a number R of blocks of increased size $N_i \times N_i, i = 1, \dots, R$ and we apply 2D DCT separately to each of these blocks, represented as $P_i \in \mathbb{R}^{N_i \times N_i}$ for size i where $i = 1, 2, \dots, R$; c(k)= $1/\sqrt{2}$ if k = 0, and 1 otherwise.

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

Thus we obtain R sets of $N_i \times N_i$ DCT coefficients $\{F_1(u,v)\}_{u,v=0}^{N_1-1}, \{F_2(u,v)\}_{u,v=0}^{N_2-1}, \dots, \{F_R(u,v)\}_{u,v=0}^{N_R-1}$. We consider their absolute value, as in Eq. 1, where u,v=0,1, ...,N$_i$-1 and N_i^2 constitutes to the number of pixels in the block of size i.

B. Coefficient Elimination

We want to discard irrelevant coefficients in each scale *i.e.* block size in order to provide robustness to photometrical distortions and to use only the informative part of the coefficients set. We discard the DC term, $\{F_i(0,0)\}_{i=1}^{R}$ in all scales to avoid 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 (Γ_i is a constant, $\Gamma_i < N_i \times N_i$) number of coefficients from the ordered list and discard the remaining ones. In this study we decide the values of Γ_i from scales $i = 1$ to $i = R$ in approximately linearly increased quantities to consider energy compactness. Thus, letting Z_i^j denotes the *j*th DCT coefficient in the ordered list of scale i, the final set of coefficients for that particular scale becomes $f_i = [Z_i^1, Z_i^2, \dots Z_i^{\Gamma_i}]$.

C. Binarization

Finally, we binarize the selected f_i DCT coefficients by *mean quantization*. Here, mean quantization is preferred since it is sensitive to outliers, so less number of high valued coefficients that are probably more discriminative pass the thresholding test. Assume that $\{\mu_1, \dots, \mu_R\}$ are the mean values of the selected DCT coefficients of scales 1 to R, then the final binary descriptor computed for each scale i will be as in Eq. 2 where $= 1, 2, \dots, \Gamma_i$.

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

The final binary descriptor for a given keypoint is obtained by concatenation of binarized DCT coefficient sets at each scale $[b_1^{1:\Gamma_1}\ b_2^{1:\Gamma_2} \dots b_R^{1:\Gamma_R}]$.

III. Experimental Setup

Oxford dataset. We first evaluated the proposed descriptor on the Oxford[1] dataset to demonstrate robustness of MB-DCT under the transformations of blur, illumination changes, viewpoints changes, and jpeg compression. 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 same border region. For the border size, we used 64 pixels which is the half of the highest block size (N$_R$) in computation of MB-DCT. We use the

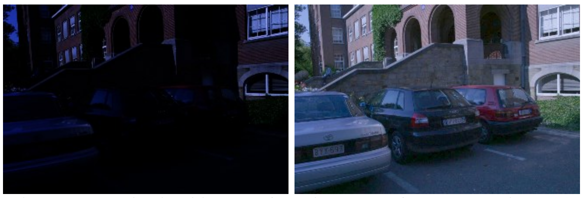

Figure 2. Synthesized images from image 6 of "Leuven" class; (a) squared brightness change, (b) square rooted brightness changes.

[1] http://www.robots.ox.ac.uk/~vgg/research/affine/

evaluation metric in [5], that we first 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 and vice versa. If the matched points are at the correct locations (with a tolerance of 2 pixels) we call them as "correct matches". We finally compute the recognition rate as (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 also evaluate MB-DCT for object recognition by implementing a Bag-of-Words type encoding on COIL-20 "processed" corpus[2], which contains 20 object categories each having 72 images with a 5 degree pose interval between. The images are in 128×128 pixels, and symmetric padding is applied with the same border size used in the previous experiment to keep 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 [15] is followed, that is, the images of each object category with pose interval of 15 degrees are taken into the training set and the remaining ones into the testing set. By computing binary descriptors densely on the images, we randomly sample a subset from the training set to input to the clustering algorithm computing the visual dictionary. The clustering algorithm is fed with the descriptors computed on the same chosen subset of training set points for all methods to create fair conditions. We use K-means as the clustering algorithm computing a visual dictionary of size 512 by using Hamming distance in similarity computation. In order to make evaluation for the existence of photometric distortions, we use original images in training set and apply distortions in Table 1 synthetically on the testing images. An example image with applied distortions is presented in Figure 3.

TABLE I. PHOTOMETRIC DISTORTIONS ON TESTING IMAGES

(b) Blurring: 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.	

IV. EXPERIMENTS

We evaluated MB-DCT in two lengths computed for R = 6 number of scales, one in 256 bits named as MB-DCT-256, and the other in 192 bits named as MB-DCT-192. $\Gamma_{\text{MBDCT-256}}=\{6,16,32,48,64,90\}$ 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

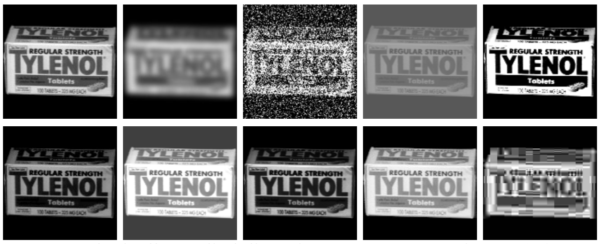

Figure 3. Synthetically applied distortions on a COIL-20 image (*Please see Table 1 for further information*). (a) Original image, (b) Blurring: PSNR=17.7 SSIM=0.5 (c) AWGN: PSNR=10.2 SSIM=0.2 (d) Contrast decr.: PSNR=11.9,SSIM=0.4 (e) Contrast incr.: PSNR=15.1,SSIM=0.9 (f) (Lin.) Brigh. Decr.: PSNR=14.7 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

implemented Bi-DCT in Matlab by the default parameters given in [11] and we executed BRIEF-256 (bits) in OPENCV library with the default parameters to make a comparative study.

A. 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 increased blur distortion even with lower descriptor length than BRIEF-256. While for the brightness changes and synthesized Leuven sequences, MB-DCT gives comparative results with BRIEF-256, BRIEF-256 outperforms MB-DCT for viewpoint changes.

B. Object recognition on Coil-20 Dataset

In this experiment we aim to explore the behaviour of the methods for object recognition task when test images were under significant amount of distortions. It should be noticed that filtering is not applied before computing the proposed MB-DCT descriptor and Bi-DCT [11], 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.

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 propose a binary descriptor called MB-DCT which proves to highly robust to photometric changes and distortions. We evaluated the performance of the proposed descriptor using two different feature sizes, on Oxford dataset and COIL20 object recognition datasets. We demonstrated that the proposed method is highly robust to blur and noise artefacts and gives comparative results when complex brightness changes

[2] http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php

exist with BRIEF even when using a shorter descriptor length of. Proposed method is not highly robust to geometrical transformations such as rotation and viewpoint changes, however we aim to improve the invariance of it to the geometric transformations as the future work.

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 [11]**
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) Br. decr.	99.79	99.79	**99.90**	85.15
(Linear) Br. inc.	**99.79**	92.50	99.69	81.77
(Nonlin.) Sq. Br. Ch.	99.79	99.58	**100**	99.69
(Nonlin.) Sq.root. Br. Ch.	99.79	99.69	**100**	99.79
JPEG compression	**99.48**	98.85	77.19	55.94

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] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint", Proc. of Computer Vision and Pattern Recognition (CVPR 2012), pp. 510-517, 2012.

[9] G. Levi and T. Hassner, "LATCH: Learned Arrangements of Three Patch Codes", arXiv preprint arXiv: 1501.03719, 15 Jan. 2015.

[10] 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.

[11] 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.

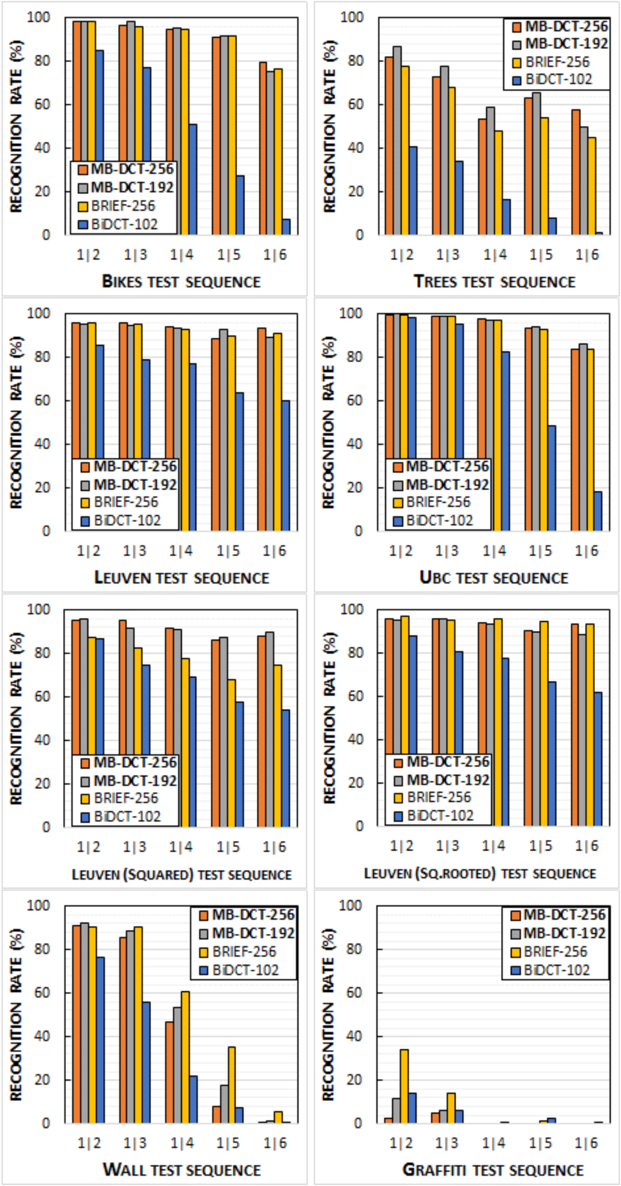

Figure 4. Recognition rates in existence of distortions

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

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

[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. 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.