Global Binary Patterns: A Novel Shape Descriptor

Erdal Sivri and Sinan Kalkan
Dept. of Computer Engineering
Middle East Technical University, Turkey
{erdal, skalkan}@ceng.metu.edu.tr

Abstract

In this paper, we propose a novel shape descriptor, called Global Binary Patterns (GBP), based on interpreting intensity values along a direction in an image as binary numbers, converting these binary numbers to their decimal values and concatenating these decimal values as the elements of a vector that is the GBP representation of the shape. Comparing with some widely-used state-of-the-art methods in the literature, we show that GBP is very fast and its performance on several widely-used databases is comparable or better.

1 Introduction

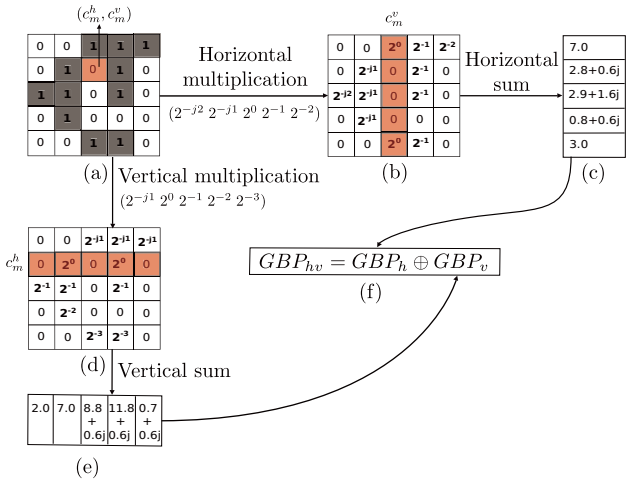


Figure 1. GBP computation. (a) The original binary image. (b) After rows are multiplied by: $(2^{-j2}\ 2^{-j1}\ 2^{0}\ 2^{-1}\ 2^{-2})$. (c) After each row is summed horizontally. (d) After columns are multiplied by: $(2^{-j1}\ 2^{0}\ 2^{-1}\ 2^{-2}\ 2^{-3})$. (e) After columns are summed vertically. (f) Resulting GBP_{hv} descriptor.

In this article, we propose a novel shape descriptor called "Global Binary Patterns" (GBP), which is a computationally simple algorithm that describes shapes using projections of binary pixel values along horizontal, vertical, diagonal and principal directions (see Fig. 1 for an illustration) - in fact, GBP can be defined for any direction. In this method, sequences of binary values, *i.e.*, bits, are interpreted as a single large binary number and converted to a decimal value. This idea is similar to the one adopted by the popular texture descriptor called Local Binary Patterns (LBP) [14] but different in that GBPs are global and defined along directions. Our comparison on several databases reveals that this simple descriptor performs comparatively against several widely used methods (LBP, Shape Context (SC) [1], Histogram of Oriented Gradients (HOG) [5], Zernike Moments (ZM) [17] and Fourier Descriptors (FD) [18]).

1.1 Related Studies and Background

Shape representation and description methods are generally analyzed in two broad categories: *contour-based* and *region-based* methods (see, e.g., [3, 19]). Some shapes are represented better with a region-based descriptor while some are better suited to a contour-based description. Therefore, in this article, we compare GBP against methods from both categories.

Popular contour-based methods include *shape signatures*, *curvature scale space*, *Fourier descriptors* and *shape context*. Shape signatures describe the shape as a one-dimensional feature vector constructed from shape boundary, using central distance, tangent angle, cumulative angle, curvature, complex coordinates and centroid distance [6]. Curvature scale space descriptors, on the other hand, utilize scale space representation of the shape contour to describe shapes [12]. Another method, Fourier descriptors, uses the frequency amplitudes of the Fourier Transform of the boundary pixels. Shape Context, represents a shape by a histogram of relative spatial configurations of pixels.

Popular region-based descriptors include *geometric moments*, *shape matrix*, *convex hull* and *axis-based descriptors*. Geometric moments [10] are based on the idea that an image can be represented as a set of moments, where a moment of an image is a particular weighted average of a function of pixel intensities. Orthogonal moments, Zernike moments, Legendre moments, rotational moments and complex moments are some of the well-known image moments used in the literature [4, 7, 17]. Shape matrix is another region-based descriptor [9], which transforms the shape into a matrix by polar quantization. In contrast to raster sampling, instead of a regular square grid, a polar raster of concentric circles are used to describe the shape. Convex hull descriptors [8] represent the shape as a string of concavities or concavity tree. Concavities of the shape are computed using the convex hull of the shape, which is defined to be the smallest convex region which covers the shape. The difference between the convex hull and region is known as the concavity of the shape. Finally, axis-based models [2] capture the interior region of the shape as a graph and are generally insensitive to articulations and occlusions.

Although they are not shape-based methods, we compare our GBP descriptor against LBP and HOG since they are widely-used simple texture-based descriptors. HOG descriptor is a histogram of gradient occurrences in localized grid cells. HOG has been successfully applied in several tracking applications, espe-

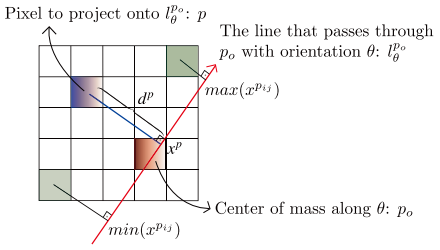


Figure 2. Illustration of the projection of a pixel onto a line that passes through the pixel p_o with orientation θ.

cially that of humans [5]. LBP, albeit computationally simple, is a powerful texture descriptor [14]. The basic idea in LBP is to compute a bit string in local windows by thresholding pixels in the window with respect to the center pixel. The bit string is interpreted as a binary number and converted to a decimal number, which is used to build a histogram.

2 Global Binary Patterns (GBP)

GBP creates a set of bit strings for any direction of a binary image and interprets these bit strings as binary numbers to build a global descriptor (see Fig. 1). The idea of extracting bit strings from intensity values is also essential in LBP; the difference is that LBP extracts these bit strings in local windows whereas GBP extracts them on global directed lines. In its simplest form, GBP of a row, r, of a binary image I is defined as follows:

$$GBP_h(r) = \sum_{c=0}^{C-c_m^h} I(r, c+c_m^h)2^{-c} + \sum_{c=0}^{c_m^h} I(r, c_m^h - c)2^{-jc}, \quad (1)$$

where c_m^h is the center of mass along the horizontal direction, and C is the number of columns in I. GBP_h computes GBP along the horizontal direction. Similarly, GBP along the vertical direction, denoted GBP_v, is defined as follows:

$$GBP_v(c) = \sum_{r=0}^{R-c_m^v} I(r+c_m^v, c)2^{-r} + \sum_{r=0}^{c_m^v} I(c_m^v - r, c)2^{-jr}, \quad (2)$$

where c_m^v is the center of mass along the vertical direction and R is the number of rows in I. See Fig. 1 for an illustration of GBP computation.

Although GBP_h and GBP_v are defined along horizontal and vertical directions, GBP can be constructed along any arbitrary direction, which may effect (as investigated in Section 3) the performance of the descriptor. To this end, we extend the definitions in Eq. 1 and 2 to incorporate *projection* along an arbitrary direction with orientation θ (see Fig. 2). Let $l_\theta^{p_o}$ be the line, with orientation θ, that passes through p_o, which is the center of mass of image I along θ. GBP_θ for image I is then defined as follows:

$$GBP_\theta(k) = \sum_{p \in I^-} \delta\left(L\gamma - k\right) I(p) 2^{-j d^p} + \sum_{p \in I^+} \delta\left(L\gamma - k\right) I(p) 2^{-d^p}, \quad (3)$$

where I^- and I^+ are the set of pixels on the negative and positive side of the center of mass respectively, d^p is the point-to-line distance between the pixel p and the line $l_\theta^{p_o}$, x^p is the projection of the pixel p onto the line $l_\theta^{p_o}$, L is the desired length of the GBP_θ descriptor (L is taken to be 32 for the experiments conducted in this paper), γ is $\frac{x^p - min(x^{p_{ij}})}{max(x^{p_{ij}}) - min(x^{p_{ij}})}$, and $\delta(\cdot)$ is the Kronecker delta defined as:

$$\delta(x) = \begin{cases} 1 & if\ x = 0, \\ 0 & otherwise. \end{cases} \quad (4)$$

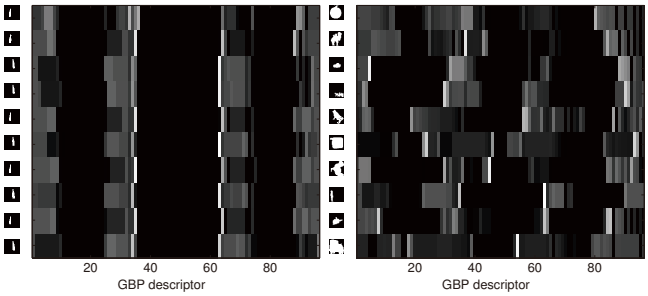


Figure 3. Sample images from the MPEG-7 database and the corresponding GBP descriptors (brighter intensities indicate higher values). Descriptors are constructed by concatenating GBPs in Eqs. 5, 7 and 9. Left: images from a single category. Right: images from different categories.

Using Eq. 3, it is possible to use any number of projections to form the GBP descriptor. Analysis in this study is performed using combinations of horizontal, vertical, diagonal and principal directions defined as follows.

$$GBP_h(I) = GBP_0(I), \quad (5)$$
$$GBP_{h'}(I) = GBP_0^{R_h}(I), \quad (6)$$
$$GBP_v(I) = GBP_{-90}(I), \quad (7)$$
$$GBP_{v'}(I) = GBP_{-90}^{R_v}(I), \quad (8)$$
$$GBP_d(I) = GBP_{-45}(I), \quad (9)$$
$$GBP_p(I) = GBP_\phi(I), \quad (10)$$

where R_h and R_v denote the reverse of the image in horizontal and vertical directions, ϕ is the orientation of the principal axis of the shape, which is computed using Principal Component Analysis (PCA) [15]. Based on the nature of the problem, different GBPs as defined in Eqs. 5, 6, 7, 8, 9 and 10 can be concatenated as a single GBP descriptor.

Fig. 3 displays GBP descriptors for sample images from the MPEG-7 database [11]. GBP descriptors are constructed using horizontal, vertical and diagonal projections: $GBP_h \oplus GBP_v \oplus GBP_d$ ($\oplus$ concatenates two vectors). As seen from the figure, descriptors extracted from similar shapes form a more uniform distribution compared to those extracted from dissimilar shapes.

Comparison between GBP descriptors is performed using the so-called Canberra distance: $d(a,b) = \sum_{i=1}^{|a|} \frac{|a_i - b_i|}{|a_i + b_i|}$ where a and b are the GBP descriptor vectors extracted from two images.

3 Evaluation and Results

We evaluate GBP on several widely-used databases, and compare GBP's accuracy and running time performances with those of some widely-used descriptors

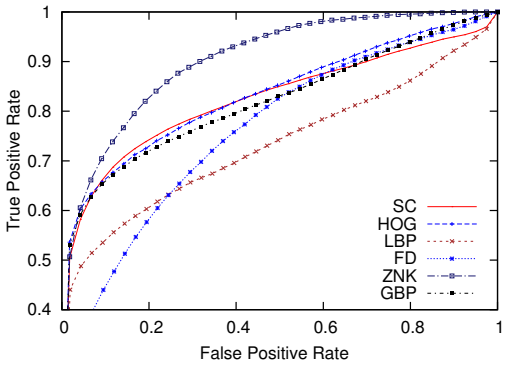
such as SC, HOG, LBP and FD. We chose these methods for comparison **because they are good representatives of shape descriptors widely used in the literature**. Databases used in the experiments are the "MPEG-7 CE Shape-1 Part-B database" [11], the "Brown University Kimia databases" [16], and the "Columbia object image library (Coil-100) database" [13] [1].

In the literature, evaluation on the MPEG-7 database is performed using the so-called *Bull's Eye* score, in which every shape is compared to all other shapes and among the top 40 similar results, the ratio of correct matches to the highest possible number of matches are reported. Kimia databases use a score called *TopRank*, in which every shape is compared to all other shapes and the number of correctly classified N nearest neighbors for each query image is reported in a tabular form. Finally, results on the Coil-100 database are reported using the accuracy measure.

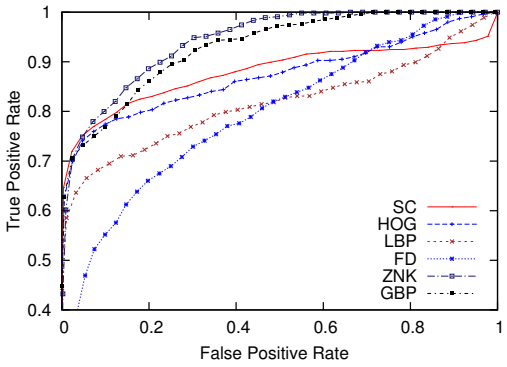
We first analyze the effect of GBP's parameters on its accuracy. GBP can be composed of combinations of several projections listed in Eqs. 5, 6, 7, 8, 9 and 10. Evaluations in this study are performed using the following GBP descriptors: GBP_d given in Eq. 9, GBP_p given in Eq. 10, and GBP_{hv} and $GBP_{hv'}$ defined as follows: $GBP_{hv} = GBP_h \oplus GBP_v$ and $GBP_{hv'} = GBP_h \oplus GBP_v \oplus GBP_{h'} \oplus GBP_{v'}$. We have also tested whether extracting GBP from equally divided grids performs better. For this, we have divided the test images into $n \times n$ many equally sized grids, where we picked up n as 1, 2 and 4. Our evaluation in the MPEG-7, Kimia-99, Coil-100 and Kimia-216 databases (detailed results not provided here for the sake of space) revealed that $GBP_{hv'}$ performed best (in terms of the metrics used for each database (see above)), regardless of the number of grids the image is divided into.

In Fig. 4, we present the retrieval results of all methods using Receiver Operating Characteristics (ROC) curves. As the figure suggests, GBP outperforms other methods on the MPEG-7 database and performs comparable to other methods on other databases (see also Fig. 5). We observe that GBP performs comparable with ZM on shape-based databases, and on par or better than HOG on the texture-based Coil-100 database.

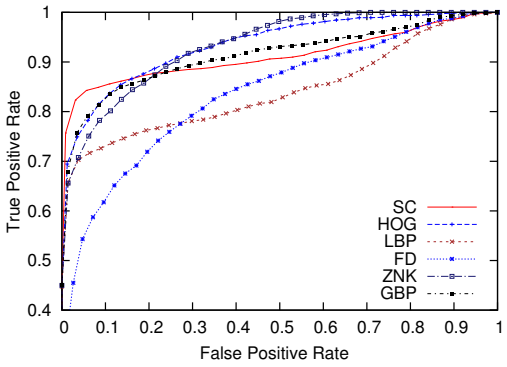
Running time comparisons are performed on Kimia-216 (arbitrarily chosen) by running each method (with their best parameter settings) in Matlab on the same desktop machine. For better evaluation, running times are recorded for feature extraction (E) and matching (M) separately. The M phase, includes an *all-to-all* matching, which uses distance metrics specific to each method. In Table 2, we see that the E phase of GBP has the best running time performance (0.12s), and the M phase of SC has the worst performance (27520.04s).



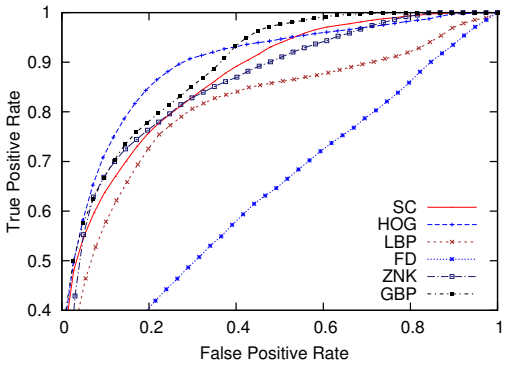
(a) MPEG-7 database



(b) Kimia-99 database



(c) Kimia-216 database



(d) Coil-100 database

Figure 4. ROC Curves for all methods with best parameter settings. For the sake of visibility, only $tpr > 0.4$ is shown without loss of any information.

4 Conclusion and Future Work

In this paper, we proposed a novel shape descriptor named Global Binary Patterns, which is similar to the popular texture descriptor, LBP. Running-time and image-retrieval performances of the method are presented in comparison with the well-known widely-used descriptors such as SC, HOG, LBP, FD and ZM on several widely-used databases. We report that GBP stands out as a promising feature descriptor since it is fast, simple and provides comparable performance to the methods analyzed in the article. For these reasons,

[1] Images are resized to 32×32 for all evaluations.

Table 1. Scores of all methods with best parameter settings.

(a) MPEG-7 database

Method	Bull's Eye
GBP	59.38
SC	58.97
HOG	57.38
ZNK	54.75
LBP	50.38
FD	28.55

(b) Coil-100 database

Method	Accuracy
HOG	99.40
SC	94.20
LBP	94.00
GBP	90.60
ZNK	89.80
FD	47.20

(c) Kimia-99 database TopRank scores

Method	1st	2nd	3rd	4th	5th	6th	7th	8th
SC	98	93	88	80	78	82	73	77
HOG	96	88	89	87	80	70	69	64
GBP	95	85	84	79	76	73	68	66
ZNK	90	85	86	81	75	71	63	67
LBP	97	85	85	75	72	65	61	57
FD	72	61	59	54	41	37	35	35

(d) Kimia-216 database TopRank scores

Method	1st	2nd	3rd	4th	5th	6th	7th	8th
HOG	209	205	201	195	194	188	180	172
SC	206	203	190	188	181	176	170	168
GBP	205	192	183	179	179	163	163	163
ZNK	197	188	182	170	164	154	148	148
LBP	199	191	179	169	158	152	153	147
FD	156	139	128	125	107	103	87	86

Table 2. Running times (sec).

Method	Extraction	Matching
SC	26.91	27520.04
HOG	0.57	6.74
LBP	18.07	6.57
FD	1.52	5.91
ZNK	1.14	4.19
GBP	**0.12**	**5.95**

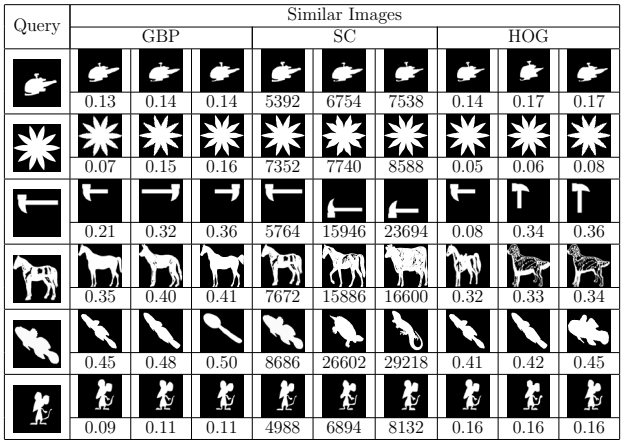


Figure 5. Query results for randomly selected images from the MPEG-7 database.

GBP is especially suitable for real-time applications or quick development requirements.

Acknowledgement

This work is partially funded by TUBITAK Project No 7110393.

References

[1] Serge Belongie, Greg Mori, and Jitendra Malik. Matching with shape contexts. In *Proceedings of the IEEE Workshop on Content-based access of Image and Video-Libraries*, 2000.

[2] H. Blum. Biological shape and visual science. *Journal of Theoretical Biology*, 38(2):205–287, 1973.

[3] M. Bober. Mpeg-7 visual shape descriptors. *IEEE Tr. on Circuits and Systems*, 1(6):716–719, 2001.

[4] J. F. Boyce and W. J. Hossack. Moment invariants for pattern recognition. *Pattern Recognition Letters*, 1(5-6):451–456, 1983.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.

[6] E. R. Davies. *Machine Vision: Theory, Algorithms, Practicalities*. Morgan Kaufmann Publishers Inc., 2004.

[7] V. P. Dinesh Kumar and T. Tessamma. Performance study of an improved legendre moment descriptor as region-based shape descriptor. *Pattern Recognition and Image Analysis*, 18(1):23–29, 2008.

[8] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., 2nd edition, 2001.

[9] A. Goshtasby. Description and discrimination of planar shapes using shape matrices. *PAMI*, 7(6):738–743, 1985.

[10] M. K. Hu. Visual pattern recognition by moment invariants. *IRE T. on Information Theory*, IT-8:179–187, 1962.

[11] L. J. Latecki, R. Lakämper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *CVPR*, 2000.

[12] F. Mokhtarian and A. K. Mackworth. A theory of multiscale, curvature-based shape representation for planar curves. *PAMI*, 14(8):789–805, 1992.

[13] S. Nene, S. Nayar, and H. Murase. Columbia object image library (coil-100). Technical report, Columbia University, 1996.

[14] T. Ojala, M. Pietikinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.

[15] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.

[16] T. B. Sebastian, P. N. Klein, and B. B. Kimia. Recognition of shapes by editing shock graphs. In *ICCV*, 2001.

[17] C. H. Teh and R. T. Chin. On image analysis by the methods of moments. *PAMI*, 10(4):496–513, 1988.

[18] D. Zhang and G. Lu. A comparative study of fourier descriptors for shape representation and retrieval. In *ACCV*, pages 646–651, 2002.

[19] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.