Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 58 (2015) 714 – 722

Procedia Computer Science

Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Discrete Cosine Transformation and Height Functions based Shape Representation and Classification

B.H.Shekar[a], Bharathi Pilar[b,*]

[a]Department of Computer Science, Mangalore University, Mangalore, Karnataka, India. Email: bhshekar@gmail.com
[b] Department of Computer Science, University College Mangalore, Karnataka, India. Email: bharathi.pilar@gmail.com

Abstract

In this paper, we propose a combined classifier model based on two dimensional discrete cosine transform (2D-DCT) and Height Functions (HF) for accurate shape representation and classification. The DCT is capable of capturing the region information and Height Functions are insensitive to geometric transformations and nonlinear deformations. The Euclidean distance metric in case of DCT and Dynamic Programming (DP) in case of HF were respectively employed to obtain similarity values and hence fused to classify given query shape based on minimum similarity value. The experiments are conducted on publicly available shape datasets namely MPEG-7, Kimia-99, Kimia-216, Myth and Tools-2D and the results are presented by means of bull's eye score and precision-recall metric. The comparative study is also provided with the well known approaches to exhibit the retrieval accuracy of the proposed approach. The experimental results demonstrate that the proposed approach yields significant improvement over some of the well known algorithms.

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Keywords: Discrete Cosine Transformation, Height Functions, Dynamic programming, Euclidean distance, Decision fusion, Combined classifier, Shape representation, Shape classification

1. Introduction

In the field of computer vision and machine learning, the task of identifying objects in an image or a video sequence is one of the fundamental problems. The identification of objects need a powerful representation, extracting the most discriminating features and also needs a suitable metric to match these features and classifying objects into an appropriate class. In most of the imaging applications, the image analysis can be reduced to the analysis of only shapes, as shape of the object contains perceptual information which is used to describe both object boundary as well as content. Shape based methods may take contour information and/or region information of the shape and it may represent either by extracting local and/or global features. Since object representations exist in different forms, it

* Corresponding author
E-mail address: bharathi.pilar@gmail.com

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)
doi:10.1016/j.procs.2015.08.092

might be too ambitious to expect a single automated representational system to cope with variety of representations and visual forms from recognition perception. Hence, we are motivated to develop a combined classifier approach that takes advantage of more than one representation scheme and emerge as a powerful classifier system.

A brief review of some well known approaches are given here. The Shape Context (SC)[1] captures the spatial distribution of every sample point relative to all other sample point on shape contour. Ling and Jacobs[2] proposed a method called Inner Distance Shape Context (IDSC), where inner distance is used to find the distance between two sample points, for the purpose of shape representation[2]. The contour points distribution histogram (CPDH)[3] descriptor computes the deformable potential at every point along a curve and makes use of EMD distance as the similarity measure for shape matching. The Height Functions proposed by Wang et al.,[4] represents the object contour by a fixed number of sample points. Each sample point is associated with a height function. There are methods where multiple features are fused that results in feature level fusion or decision level fusion for the purpose of better retrieval accuracy. One such method is learning manifold approach[5], in which dissimilarity matrices of two categories of shape descriptors are fused resulting in considerable improvement in retrieval results. For every shape sample, centroid distance, farthest distance, Zernike distance and major axis shape descriptors are obtained. Then the dissimilarity matrices have been fused and the adjacency matrix is formed for the purpose of manifold learning. On the similar line, Local Binary Pattern and Height Functions[6], Inner distance Shape context and Local Binary Pattern[7] contributions have been explored for shape representation and classification. In the context of designing combined classifier, we have made an attempt to integrate 2D-DCT that capture spatial distribution and Height Functions that captures the contour characteristics and hence to achieve better classification accuracy.
The remaining part of the paper is organized as follows. In Section 2, the technique of Discrete Cosine Transformation (DCT) is presented. In Section 3, an insight into Height Functions is given. The proposed approach is given in Section 4. In Section 5, the experimental set-up along with results are brought out and conclusion is given in Section 6.

2. Discrete Cosine Transformation

The Discrete cosine transform (DCT)[8] has been widely used in the domain of Image Processing and Pattern Recognition. DCT extracts the frequency domain information from the given object and represents object in terms of few set of coefficients. In DCT coefficients matrix, the first coefficient is called DC coefficient and the other coefficients are referred to as AC coefficients. The frequency of the coefficients increases from left to right and from top to bottom. The DCT coefficients with larger magnitude are mainly concentrated at the upper-left corner which represents the main components of the spectral coefficients of the image. This low frequency part carries most of the visually significant information of the image data. The coefficients at the lower right corner of the matrix represent high frequency part with small amplitude. Since the main energy of the image is concentrated in the low frequency, we can discard some of the AC coefficient values which are close to or equal to 0. For an M x N image, the 2D DCT is given by:

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) \times cos\frac{\pi(2x+1)u}{2M} cos\frac{\pi(2y+1)v}{2N} \tag{1}$$

where,

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{M}}; u = 0 \\ \sqrt{\frac{2}{M}}; u = 1, 2 \ldots M-1 \end{cases} \tag{2}$$

and, x and y are spatial coordinates in the image, and u and v are coordinates in the transformed image. The 2D-DCT coefficients are read in a zigzag manner starting from the top-left corner of the DCT matrix and can be converted to a one dimensional vector. A typical 10x10 DCT matrix transform will have most of the energy relocated to its upper-left corner with the DC coefficient, F_{00} representing the proportional average of the sample values of the DCT matrix. The AC coefficients of the DCT matrix represents the change in the intensity among the samples.

3. Height Functions

The Height Functions (HF) is one of the well known contour based shape descriptor proposed by Wang et al.,[4], which is based on the work of Liu et al.,[9]. In this approach, a fixed number of sample points on shape contour is considered in order to compute the height functions of a sample point, say x and the height function for x is defined as the distance of other sample points to its tangent line. To illustrate, let $p_1, p_2 \dots p_n$ be the n sample points on shape contour in a counter clockwise direction. The first step in computing the height functions is to identify a reference axis, which is the tangent line, say l_i, for every point p_i. Let $h_{i,k}$ be the distance between k^{th} sample point to tangent line l_i. The height functions are signed in nature. The sign is positive, negative or zero depending on whether k^{th} sample point is to the left, right or on the tangent line l_i, thus providing the information about both the distance and the location of the sample point with respect to the tangent line.

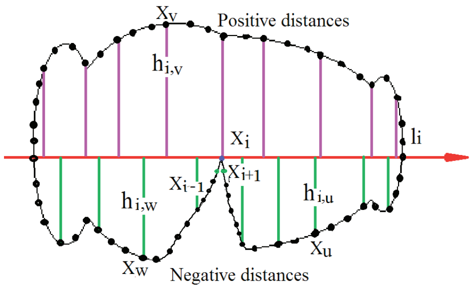

Fig. 1. Height Functions of the jar shape of MPEG-7 dataset

The Figure 1 illustrate the height function for a sample point x_i. The height value $h_{i,v}$ is positive since sample point x_v is to the left of the axis l_i and $h_{i,u}$is negative since the location of x_u is to right of the axis. Thus, height values are calculated for every sample points on the contour and are of the same order as the sequence of sample points considered. The height values $h_{i,k}(k = 1, \dots, n)$ of the k^{th} sample point x_k with respect to axis l_i of the point x_i is defined as

$$\begin{aligned} H_i &= (h_i^1, h_i^2, ..., h_i^n)^T \\ &= (h_{i,i}, h_{i,i+1}, ..., h_{i,n}, h_{i,1}, h_{i,2}, ..., h_{i,i-1})^T \end{aligned} \tag{3}$$

It is observed that h_i^1=$h_{i,i}$= 0 for every i = $1, ..., n$. The height functions defined above are insensitive to geometric transformations. It is found that HF is sensitive to local contour deformations and hence to overcome this problem smoothing function is designed which is defined as follows.
For a given integer k, the sequence of integers $1, 2, \dots, n$ are divided into disjoint intervals $[1, k]$, $[k + 1, 2k]$, . . ., and the average value of the height values in each interval is computed.

$$f_i^j = \frac{1}{k} \sum_{t=(j-1)k+1}^{jk} h_i^t, \tag{4}$$

where j = $1, ..., m$ with m = $\lfloor n/k \rfloor$ and the arithmetic is modulo n. The smoothing process also reduces dimensionality from n to m by the ratio k, since $k > 1$ and $m < n$ consequently. In order to make shape representation scale invariant, smoothed height values are subjected to normalization by maximal absolute value. The matching of the test shape and training shape is achieved by means of dynamic programming method. The distance between two shapes is computed by finding the optimal correspondence of contour points and the dissimilarity value is the sum of the distances of these corresponding points. This dissimilarity value is used to rank the database of shapes for shape retrieval.

4. Proposed Approach

In our work, we proposed a combined classifier model that integrates 2D-Discrete Cosine Transformation and Height Functions to represent and classify shapes accurately. The Discrete Cosine Transformation, which captures the spatial information of the shape and the height functions corresponding to sample points of the shape contour are calculated. The Euclidean distance measure is used to compare the DCT values and Dynamic programming (DP) algorithm is used for shape matching in case of height functions.

4.1. Two dimensional Discrete Cosine Transform

The feature extraction is carried out using two dimensional discrete cosine transform as follows

1. For every shape apply 2D-DCT method and obtain the transformed DCT matrix.
2. Take only 10 x 10 matrix of the top left corner of DCT matrix.
3. Convert the above 2D matrix into a 1D array applying zigzag scan
4. Normalize the vector and store in the knowledge base.

Repeat the above process for every shape in the training set forming the DCT based knowledge base.

4.2. Feature extraction based on Height Function

The height functions of the given shape are calculated as follows

1. Extract the contour of the given shape.
2. Select N sample points along the contour in counter-clockwise direction, (N=100 in our experiment).
3. Compute height function vectors for each sample points (Eq. (3)).
4. Obtain the smoothed height function vectors (Eq. (4)).
5. Normalize the smoothed height functions by dividing them with maximal absolute value.

The above process is repeated for all the shapes in the training set to form the height functions based knowledge base. Thus every shape is described by 2D-DCT and HF features forming a hybrid knowledge base.

4.3. Classification

Given the query shape, the classification is done as follows.

1. Compute the 2D-DCT matrix of the test shape and obtain its one dimensional vector applying zigzag scan, say D_t.
2. Extract the height functions based feature vectors of the test shape say H_t
3. In case of DCT feature vectors, compute the distance between the test shape with the sample shapes of the training set using Euclidean distance measure.
4. In case of height functions vectors, Dynamic programming (DP) algorithm is used to find out the correspondence between contour sample points of two shapes. Let D_T and D_H be the distances obtained due to DP and Euclidean distance measure.
5. The distance vectors D_T and D_H are fused to form the resultant vector D_R as follows

$$D_R = D_T + \beta D_H \tag{5}$$

   where, the value β is obtained experimentally, which are variant to the dataset
6. The distance values are arranged in the ascending order of the distance to obtain top shape matches.

5. Experimental Results and Discussions

In this section, we present the experimental results on the standard shape datasets namely: *MPEG-7*, *Kimia-99*, *Kimia-216*, *Myth* and *Tools-2D*. The performance of the proposed approach is demonstrated through *Bull's eye score*[2]. In addition, we have also presented retrieval rate depicting top-n closest matching shape.

5.1. Experimental Results on MPEG-7

The MPEG-7 dataset consists of 1400 shapes from 70 classes with each class consisting of 20 shape samples. The retrieval rate obtained due to proposed approach for each shape class considering top 40 retrievals is shown in Figure 2. The retrieval rate of the proposed approach on MPEG-7 is calculated and the results obtained due to proposed approach along with the results of the state-of-art algorithms is tabulated in Table 1. We notice that the proposed method achieves the better score of **91** percent on MPEG-7 data set. The retrieval rate depicting top 12 closest matching shape is given Table 2. It shall be observed from Figure 2 that more than 50 percent of the classes exhibit almost 100 percent accurate results, where as shapes belong to device classes exhibit little low performance due to high intra-class dissimilarities and high inter class similarities.

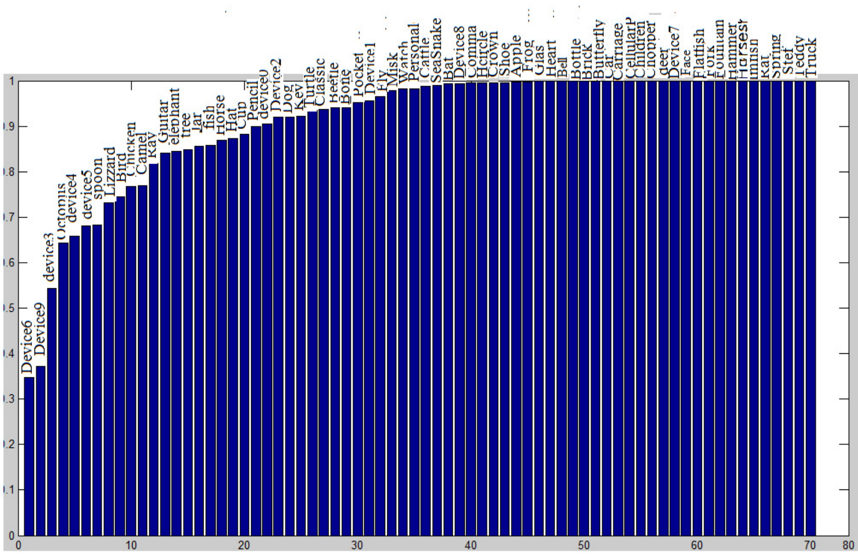

Fig. 2. Class-wise retrieval results for MPEG-7 dataset.

In addition, we have also shown in Figure 3 the improvements made by the proposed approach by presenting the top-12 retrieved shapes for a given query shape and compared with the retrieval results of Height Function approach [4].

The graphs of precision and recall for Height Functions[4] alone and proposed approach are given in Fig 4. The proposed approach exhibit better precision and recall rate.

5.2. Experimental Results on Kimia-99

The Kimia-99 shape dataset consists of 9 classes with 11 samples in each class. The 99 shapes belong to 9 different classes. The retrieval rate depicting the top-10 matching shapes is presented in Table 3. In addition, we have also presented the retrieval results on Kimia-99 database considering some of the well known shape representation techniques. It shall be observed from Table 3 that the proposed approach exhibits better performance when compared to other approaches.

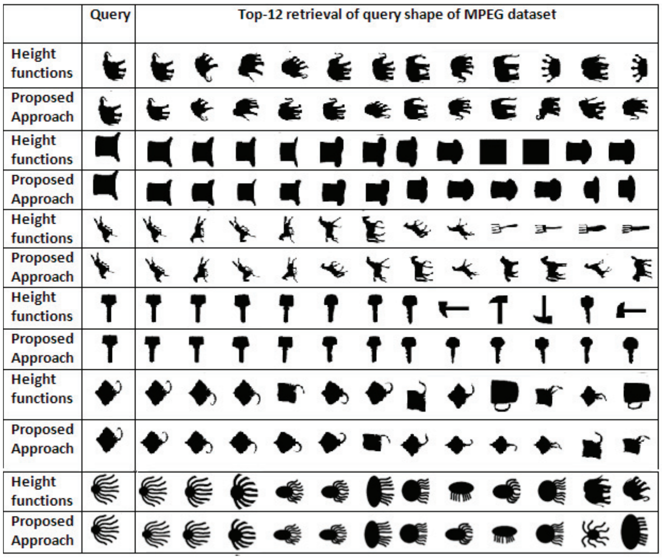

Fig. 3. Top 12 retrieved shapes of query shape by Height Functions [4] and proposed approach

Table 1. Retrieval rate(Bull's eye Score) for MPEG-7 dataset [10]- A comparative Analysis

DataSet	MPEG-7
Proposed Approach	**91.00**
HT+PS [11]	89.88
Height functions [4]	89.66
Locally affine invariant descriptors [12]	89.62
Contour flexibility [13]	89.31
Learned manifold [5]	88.52
Two strategies [14]	88.39
Aspect shape context [15]	88.30
Hierarchical parts [16]	88.30
Shape tree [17]	87.70
TAR + shape complexity + global [18]	87.23
TAR + shape complexity [19]	87.13
SC + DP [20]	86.80
HPM [21]	86.35
Symbolic representation [22]	85.92
PS+IDSC [23]	85.82
IDSC + DP [2]	85.40

Table 2. Top 12 closest matching shapes for MPEG-7 dataset

Dataset	1	2	3	4	5	6	7	8	9	10	11	12
PS+LBP [24]	1400	1345	1276	1221	1157	1113	1070	1023	995	961	933	898
IDSC [2]	1400	1375	1342	1315	1256	1235	1209	1188	1134	1108	1048	1045
HF [4]	1400	1384	1358	1352	1302	1283	1275	1251	1223	1193	1169	1153
HT+PS [11]	1400	1384	1363	1356	1305	1290	1281	1265	1226	1198	1172	1167
Proposed Approach	**1400**	**1386**	**1381**	**1365**	**1327**	**1300**	**1290**	**1279**	**1232**	**1233**	**1201**	**1193**

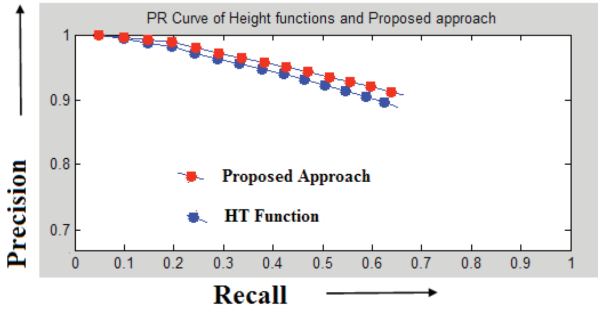

Fig. 4. Precision-Recall diagram for MPEG-7 dataset

Table 3. Top 10 closest matching shapes for Kimia's 99 dataset

Approach	1	2	3	4	5	6	7	8	9	10	Total
PS+LBP[24]	99	97	97	88	88	86	86	90	80	77	888
Learn. manifold[5]	99	99	98	98	98	96	95	89	80	65	917
Hierarch. Parts[16]	99	99	98	98	98	97	96	94	93	82	954
IDSC[2]	99	99	99	98	98	97	97	98	94	79	958
Shape tree[17]	99	99	99	99	99	99	99	97	93	86	969
Two strategies[14]	99	99	99	98	99	99	99	97	96	84	969
HF[4]	99	99	99	99	98	99	99	96	95	88	971
Symbolic Repr.[22]	99	99	99	98	99	98	98	95	96	94	975
Proposed Approach	**99**	**99**	**99**	**99**	**97**	**98**	**99**	**97**	**99**	**96**	**982**

5.3. Experimental Results on Kimia-216

The Kimia-216 shape dataset consists of 18 classes with 12 samples in each class. The top 11 closest matches obtained due to the proposed methodology is shown in Table 4. In Table 4, the retrieval results on Kimia-216 database for the proposed method and other recent well known methods are placed together. One can notice here too that the proposed approach exhibit better performance when compared to other methods.

Table 4. Top 11 closest matching shapes for Kimia-216

Approach	1	2	3	4	5	6	7	8	9	10	11	Total
SC[1]	214	209	205	197	191	178	161	144	131	101	78	1809
CPDH+EMD(Eucl)[3]	214	215	209	204	200	193	187	180	168	146	114	2030
CPDH+EMD(shift)[3]	215	215	213	205	203	204	190	180	168	154	123	2070
PS+LBP[24]	216	209	205	195	195	197	188	180	179	163	152	2079
HF[4]	216	216	216	215	215	212	211	204	200	194	179	2278
IDSC[2]	216	215	216	214	211	214	210	212	207	204	193	2312
Proposed Approach	**216**	**216**	**216**	**215**	**214**	**212**	**210**	**210**	**207**	**203**	**195**	**2314**

Table 5. Bull Eye Score(Retrieval Rate) of proposed method for Kimia 99, Kimia 216, MPEG-7 dataset

DataSet	Bull eye test
Kimia-99	100
Kimia-216	98.84
MPEG-7	91.00

The overall performance of the proposed approach measured using bull's eye score on Kimia-99, Kimia-216 and MPEG-7 shape datasets is given in Table 5.

5.4. Experimental Results on Myth dataset

Myth data set contains 15 shapes (5 humans, 5 horses and 5 centaurs). We have also presented in Table 6, the average distance within the shape classes and between the shape classes. The experiments exhibit that the combined distance measure suits well to separate the objects belong to different classes. It shall be observed from Table 6 that the intra-class distances are substantially less than the inter-class distances.

Table 6. Average distance within classes and the average distance between the classes for the Myth dataset

	centaurs	horses	people
centaurs	2.662553717	5.569424549	12.95656119
horses	5.569424549	4.007757096	14.97722919
people	12.95656119	14.97722919	3.052862898

A comparative analysis of the top-5 retrieval results of Myth database for Height Functions[4], IDSC[2] and the proposed approach is presented in the Table 7.

Table 7. Top 5 closest matching shapes for Myth dataset

Approach	1st	2nd	3rd	4th	5th	Total
IDSC[2]	15	15	9	4	10	53
Height Functions[4]	15	14	12	10	9	60
Proposed Approach	**15**	**14**	**14**	**15**	**10**	**68**

5.5. Experimental Results on Tools dataset

The Tools-2D consisting of 35 objects of various instruments (a pair of pliers, knives, scissors of different types). A comparative analysis of the top-5 retrieval results of tools database for height functions and the proposed approach is presented in the Table 8. One can notice here too that the proposed approach classification accuracy is better than the height function alone and also better than IDSC[2] which is found to be one of the best shape representation technique.

Table 8. Top 5 closest matching shapes for Tools-2D dataset

Approach	1st	2nd	3rd	4th	5th	Total
IDSC[2]	35	26	16	12	10	99
Height functions[4]	35	35	33	30	24	157
Proposed Approach	**35**	**35**	**35**	**30**	**25**	**160**

6. Conclusion

In this work, we have designed the combined classifier model for binary shapes representation based on Discrete Cosine Transformation and Height Function followed by classification using Euclidean distance and DP as the distance measure. The characteristics of height function such as rotation invariance / insensitivity to noise and occlusion along with DCT's powerful energy compaction capturing spatial information is found to be suitable choice for shape representation which is demonstrated through extensive experimentation. Experimental results on standard shape databases namely MPEG-7, Kimia-99, Kimia-216, Myth and Tools-2D datasets exhibit the success of the proposed fusion approach.

Acknowledgements

The authors would like to thank the support provided by the Department of Science and Technology, Govt. of India vide Project no. INT/RFBR/P-133 and Russian Foundation for Basic Research, Russia vide Project no. RFBR 12-07-92695.

References

1. Belongie, S., Malik, J., Puzicha, J.. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2002;**24**(4):509–522.
2. Ling, H., Jacobs, D.W.. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2007;**29**(2):286–299.
3. Shu, X., Wu, X.J.. A novel contour descriptor for 2d shape matching and its application to image retrieval. *Image and vision Computing* 2011;**29**(4):286–294.
4. Wang, J., Bai, X., You, X., Liu, W., Latecki, L.J.. Shape matching and classification using height functions. *Pattern Recognition Letters* 2012;**33**(2):134–143.
5. Chahooki, M.A.Z., Charkari, N.M.. Learning the shape manifold to improve object recognition. *Machine vision and applications* 2013; **24**(1):33–46.
6. Shekar, B.H., Pilar, B.. Shape representation and classification through height functions and local binary pattern-a decision level fusion approach. In: *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on*. IEEE; 2014, p. 2481–2488.
7. Shekar, B.H., Pilar, B., Kittler, J.. An unification of inner distance shape context and local binary pattern for shape representation and classification. In: *Proceedings of the 2nd International Conference on Perception and Machine Intelligence*. ACM; 2015, p. 46–55.
8. Er, M.J., Chen, W., Wu, S.. High-speed face recognition based on discrete cosine transform and rbf neural networks. *Neural Networks, IEEE Transactions on* 2005;**16**(3):679–691.
9. Liu, H., Latecki, L.J., Liu, W.. A unified curvature definition for regular, polygonal, and digital planar curves. *International Journal of Computer Vision* 2008;**80**(1):104–124.
10. Latecki, L.J., Lakamper, R., Eckhardt, T.. Shape descriptors for non-rigid shapes with a single closed contour. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings*; vol. 1. IEEE; 2000, p. 424–429.
11. Shekar, B.H., Pilar, B.. An ensemble of morphology based pattern spectrum and height functions for shape representation and classification. In: *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on*. IEEE; 2014, p. 2264–2272.
12. Wang, Z., Liang, M.. Locally affine invariant descriptors for shape matching and retrieval. *Signal Processing Letters, IEEE* 2010;**17**(9):803–806.
13. Xu, C., Liu, J., Tang, X.. 2d shape matching by contour flexibility. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2009; **31**(1):180–186.
14. Temlyakov, A., Munsell, B.C., Waggoner, J.W., Wang, S.. Two perceptually motivated strategies for shape classification. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2010, p. 2289–2296.
15. Ling, H., Yang, X., Latecki, L.J.. Balancing deformability and discriminability for shape matching. In: *Proceedings of the 11th European Conference on Computer Vision*; ECCV'10. Springer-Verlag; 2010, p. 411–424.
16. Payet, N., Todorovic, S.. Matching hierarchies of deformable shapes. In: *Proceedings of the 7th IAPR-TC-15 International Workshop on Graph-Based Representations in Pattern Recognition*; GbRPR '09. Springer-Verlag; 2009, p. 1–10.
17. Felzenszwalb, P.F., Schwartz, J.D.. Hierarchical matching of deformable shapes. In: *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2007, p. 1–8.
18. Alajlan, N., El Rube, I., Kamel, M.S., Freeman, G.. Shape retrieval using triangle-area representation and dynamic space warping. *Pattern Recognition* 2007;**40**(7):1911–1920.
19. Alajlan, N., Kamel, M.S., Freeman, G.H.. Geometry-based image retrieval in binary image databases. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2008;**30**(6):1003–1013.
20. Bai, X., Wang, B., Yao, C., Liu, W., Tu, Z.. Co-transduction for shape retrieval. *IEEE Transactions on Image Processing* 2012; **21**(5):2747–2757.
21. McNeill, G., Vijayakumar, S.. Hierarchical procrustes matching for shape retrieval. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; vol. 1. IEEE; 2006, p. 885–894.
22. Daliri, M.R., Torre, V.. Robust symbolic representation for shape recognition and retrieval. *Pattern Recognition* 2008;**41**(5):1782–1798.
23. Shekar, B.H., Pilar, B.. A decision level fusion of morphology based pattern spectrum and idsc for shape representation and classification. In: *Contemporary Computing and Informatics (IC3I), 2014 International Conference on*. IEEE; 2014, p. 908–914.
24. Shekar, B.H., Pilar, B.. Shape representation and classification through pattern spectrum and local binary pattern–a decision level fusion approach. In: *Fifth International Conference on Signal and Image Processing (ICSIP)*. IEEE; 2014, p. 218–224.