Journal of ELECTRICAL ENGINEERING, VOL 73(2022), NO3, 197–202

sciendo

PAPERS

Quadrant-based contour features for accelerated shape retrieval system

Mustafa Eren Yildirim

Shape representation and retrieval are essential research topics of computer vision. This paper proposes a novel feature set to be used in content-based image retrieval systems. The proposed method is an extended version of our previous study which uses contour information of shapes. The previous study calculated the center of mass (CoM) of the shape. By taking the CoM as origin, we created imaginary vectors in every angular direction. From each vector, we extracted three features which are the number of intersections between vector and contour, average distance of intersection points to CoM, and standard deviation of these points. In this method, we extract novel features and decrease the size of the feature set to decrease the computation time. We divide the shape into quadrants and represent each quadrant by nine features. Each shape image is represented by a 4x9 feature vector. We tested the proposed method on MPEG-7 and ETH-80 datasets and compared it with the state-of-art. According to the results, our method decreased the computation time dramatically while giving a state-of-art level retrieval accuracy.

K e y w o r d s: image retrieval, shape representation, contour-based feature

1 Introduction

Shape representation and retrieval are gaining more attention from researchers due to the increase in smart services in digital image and video technology. In recent years, the number of digital images in any digital platform has increased tremendously. Thus, retrieval of a certain image from a large set of images fast and accurately is strongly required [1]. Thus, the image should be represented by the minimum number of features which should not be affected by any deformation such as rotation, illumination, scale, and translation [2].

In the current literature, image retrieval methods are grouped as content-based image retrieval (CBIR) and text-based image retrieval (TBIR). The latter is an older and less used method compared to the former. Every image has some textual information such as name, tag, social platform, *etc.* TBIR systems aim to retrieve images according to the input keywords and texts describing the image [3]. This method can be successful if all the images are annotated correctly. However, TBIR can fail if there are images without any annotations in the dataset. Moreover, these annotations are user-dependent which means an image can be searched or represented by different keywords by a different person [4].

Due to the aforementioned disadvantages of TBIR, researchers introduced and have been using CBIR for a long period in various fields such as medical applications [5], biodiversity information systems [6], digital libraries [7, 8]. CBIR systems use different feature domains like color [9, 10], texture [11-14], spatial [15], shape [16] to represent the image. Shape-based methods are widely used to represent various types of shapes accurately. They are divided mainly into two groups as contour- based and region-based methods. Region-based methods take the entire pixels within a shape region into consideration to obtain the shape representation [17, 18]. On the other hand, contour-based methods focus on the boundary of the shapes in the image [19, 20].

The major steps of a contour-based shape retrieval method are dataset generation and retrieval. In dataset generation, contour-based features are extracted from each image in the dataset and stored. A preprocessing step is used if necessary. In the retrieval phase, the same features are extracted from the query image and compared with all images in the generated dataset by using a distance metric (L1, L2, bhattacharya). According to the similarity score obtained from comparison, the class or type of the shape is detected.

In our previous study [21], we introduced angle-wise contour statistics for shape representation. The method was extracting three features for each angle in range $[0, 359°]$ by taking the shape's center of mass (CoM) as the origin. The extracted features were scale and translation invariant because of the normalization step. However, they were not rotation invariant. To overcome this problem, the authors used circular shift matching to find the best match between the query image and images in the dataset. This led to very high time consumption in the retrieval stage.

This paper focuses on the time consumption problem. We extract the same features as in [21]. Next, we divide

Department of Electrical and Electronics Engineering, Bahcesehir University, Istanbul, Turkey, mustafaeren.yildirim@eng.bau.edu.tr, Department of Electronics Engineering, Kyungsung University, Busan, Republic of Korea

https://doi.org/10.2478/jee-2022-0026, Print (till 2015) ISSN 1335-3632, On-line ISSN 1339-309X

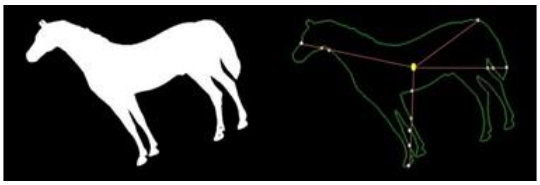

Fig. 1. Illustration of vectors for a binary image

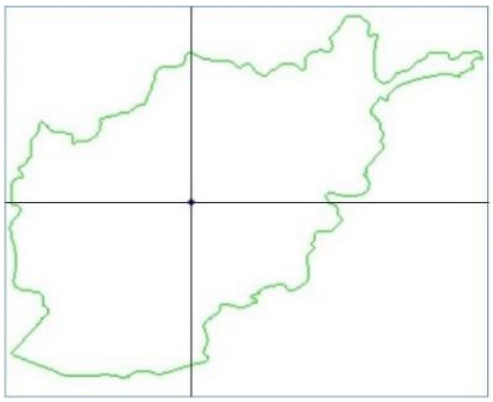

Fig. 2. Quadrants of an contour image

the shape into four quadrants. For each feature, we calculate the maximum value, standard deviation, and mean. This is repeated by each quadrant. As a result, the resulting feature space is dramatically decreased. Thus, the time spent during the shift matching is decreased. Due to the dimension reduction in feature set size, there is a very small decrease in the retrieval accuracy. However, the change in the accuracy is comparatively much smaller than the change in the computation time.

2 Related studies using contour

In this section, some of the recent studies which focused on contour-based shape retrieval are presented. Wang [22] presented a new concept called bag of contour fragments (BCF) which decomposes the contour of a shape into smaller fragments and represents each fragment by a shape descriptor. The coded fragments are then pooled to represent the entire shape. In another study [23] authors introduced a hybrid model consists of three shape descriptors which are region area descriptor (RAD), region skeleton descriptor (RSD), and simplified shape descriptor (SSD). A study [19] proposed to represent a shape by beams originating from a point on the outline to all other points on the outline. A novel shape descriptor called integral contour angle (ICA) [24] is introduced for leaf identification. There are two groups of vectors that originates from an arbitrary point on the contour of the shape. The vectors in one group go to the contour points on the right side of the origin point while the vectors of the second group go to the points on the left. The average of the vectors is merged to generate the descriptor. In [25] authors introduced a descriptor named contour-point signature (CPS). They aimed to obtain correspondences of points from the boundary of two random shapes and create a map a listed sequence of boundary points between the shapes.

3 Background and motivation

We briefly explain the background of this study. In [21], the authors defined three features by using the contour of a shape. There are 360 imaginary vectors originate from the CoM of the shape and rotating around the clock so that there is a vector traveling in the direction of each integer-valued angle. An example is shown in Fig. 1. For a binary image of horse shape, the contour is extracted. The figure illustrates only four vectors originate from CoM and travels in four different directions.

All of the vectors cut the shape contour at least one point. For each vector, three features are extracted. The first feature is the number of intersections () of the vector line and contour in that specific angle. In Fig. 1, if we start from 0 deg and rotate counterclockwise, the number of intersections are 3, 1, 3 and 5 respectively.

The second feature d_α is the average distance of the intersection points to the CoM. The last feature d_σ is the standard deviation of those distances. The last feature carries information about the distribution of the intersection points, which is a unique feature for different shapes. As a result of this process, each shape is represented by a vector with size $[360 \times 3]$. The feature order in vectors start from 0° in unit circle. Therefore, if the above procedure is applied to an image and its rotated version, the resulting vectors would be different. Manhattan distance (L^1) calculated in each feature space would be nonzero where it actually should not be. Therefore, circular shift matching is used to overcome this problem. When two vectors are compared by using the circular shift matching method, one of the vectors is stationary while the other vector is shifted by a predefined number.

Comparison is made in every shifting step by calculating the (L^1) distance for each of the three features. Retrieval is realized depending on three features individually. Instead of a feature fusion which may cause a decrease in performance, the maximum score is chosen as the retrieval score.

When the dataset and vector are small, this process does not create a high computation time. However, in an image retrieval problem where there is a mass number of images in the dataset, we need to decrease the vector size to decrease the computation time.

4 Proposed method

The proposed method decreases the feature vector size by extracting different types of features but with much less amount per feature. In [21], a shape is represented

Fig. 3. Example images of MPEG-7 part B [26,27]

by 1080 feature samples. We will decrease this number to a very small amount.

The first step of the proposed method is to divide the shape image into quadrants as shown in Fig. 2.

The quadrants are named as Q_1, Q_2, Q_3 and Q_4 which correspond to the angle ranges $[0°, 89°]$, $[90°, 179°]$, $[180°, 269°]$ and $[270°, 359°]$ respectively.

The study [21] gives out 90 numbers of each three features n_α, d_α and σ_α for each quadrant. We introduce three new features by using the features from [21]. The new features of the proposed method are the maximum value, the average value and the standard deviation of a quantity in the population of 90 samples.

The first feature (maximum value) describes the peak amount in distance, in curvature and diversity of curves.

The second feature (average value) carries more smooth information about the curvature, distance and diversity of the contour in the quadrant.

The last but not least, our third feature (standard deviation) describes the amount of change in curvature, distance to CoM and deviation among the samples.

When we apply this to each feature defined in [21], we obtain nine new features. Since normalization is made after the extraction of the initial three features, it is not necessary to normalize the new features one more time. The new features are given by

$$\left(\max_n, \max_d, \max_\sigma, \mu_n, \mu_d, \mu_\sigma, \sigma_n, \sigma_d, \sigma_\sigma\right). \tag{1}$$

The same process is done for all of the four quadrants which results with a new feature vector of $[4 \times 9]$. As a result, the amount of feature samples is decreased from 1080 to 36. The number of shifting is also decreased from 360 to 4 during image retrieval with circular shift matching which leads to a decrease in computation time. In each vector-wise matching, L^1 distance is used between features. Similar to [21], retrieval is made according to nine features separately and obtain nine different scores. The highest score is chosen as the final.

5 Experimental results

In this section, we explain the used datasets, experiment procedures, and performance results of the proposed method. All parts of the proposed method are written in Python and run on a computer with Intel(R) Core(TM) i5-10400 2.90 GHz CPU and 8 GB RAM under Windows 10 operating system.

5.1 Datasets

The proposed method is tested on two publicly available datasets namely MPEG-7 part B [28] andETH-80 [29].

MPEG-7 dataset is composed of 1400 binary images of 70 different classes as shown in Fig. 3. There are 20 images per class with variations such as scale, translation, rotation, and deformation. This dataset is generally used for image retrieval. The most common performance metric for this dataset is the bull's eye rating (BER). The first step of this metric is the bull's eye score in which every image in the dataset is used as the test image and the similarity scores of the test image with all the remaining images are calculated. The number of images in the same object class with the test image among the 40 images with the highest similarity score is called the bull's eye score. In MPEG-7 dataset, the maximum possible value of bull's eye score is 28000 since there are 1400 images with 20 images per class. The bull's eye rating is

$$r_{\mathrm{BE}} = \frac{I_{\mathrm{CR}}}{2800} \times 100, \tag{2}$$

where I_{CR} - is the number of correctly retrieved images.

The second dataset used in this study is ETH-80. This dataset includes both colored and binary versions of the same classes. Only the binary section is considered in this study. It contains 8 classes (apple, cow, car, cup, dog, horse, pear, tomato) as shown in Fig. 4. For each class, there are 10 objects which contain a large amount of intra-class variations. There are 41 images from viewpoints that are located apart from each other with

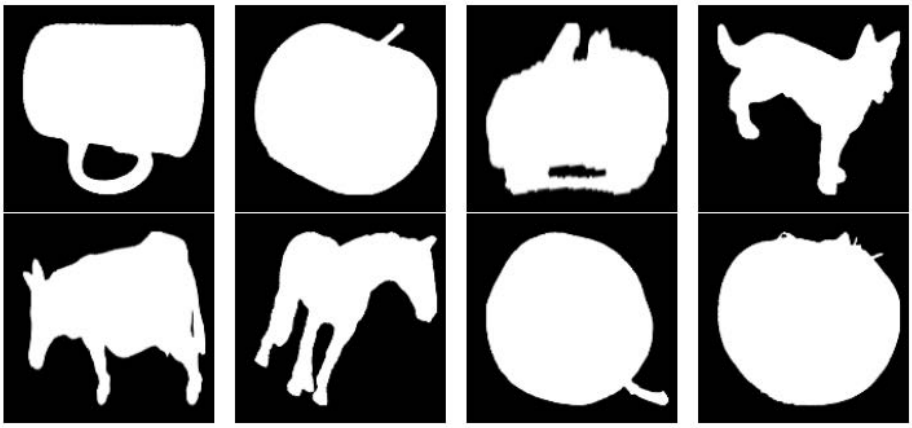

Fig. 4. The 8 categories of the eth-80 dataset

Table 1. Bull's eye rating of the state-of-the art methods for MPEG-7 Part B dataset

Method	BER(%)	TIME(ms)
Proposed method(4 pieces)	89.6	0.255
Proposed method(2 pieces)	87.9	0.158
Proposed method(1 piece)	80.44	0.103
AIR+LCDP+TSR (30)	100	2.735
IDSC+SC+SCA (31)	99.01	6.190
IDSC+GMM-r (32)	93.41	2.735
(21)	90.43	1.835
Shape vocabulary (33)	90.41	0.120
MSFDGF-SH-SF+MD (34)	87.76	0.015
MSFDGF-SH-SF+SD (34)	87.76	0.029
Shape tree (35)	87.7	6.376
HSC (36)	87.31	0.0164
TAR+DP (37)	87.13	4.150
SC+DP (38)	86.8	3.455
IDSC+DP (39)	85.4	2.735
DIR (40)	77.69	0.004
MDM (41)	70.46	0.021

Table 2. Recognition rate of the state-of-art for ETH-80 dataset

Method	RR(%)	TIME(ms)
Proposed method(4 pieces)	99.83	0.293
Proposed method(2 pieces)	94.60	0.208
Proposed method(1 piece)	94.15	0.097
(21)	99.92	2.637
MCV (42)	92.25	NA
BCF (22)	91.49	NA
Kernel-edit (43)	91.33	NA
Robust symbolic (44)	90.28	76.5
Height function (45)	88.72	NA
IDSC+DP (39)	88.11	NA

equal distances over the upper viewing hemisphere in a range (22.5°–26°), [27]. The accepted performance measure method of this dataset is leave-one- object-out cross-validation. This method is applied by taking a single object as a query object and testing it on the remaining 79 objects. The results are averaged over the entire 80 objects.

5.2 Experimental results

In this section we present the performance of the proposed method on both datasets. We compared our method with the conventional method [21] and the state-of-art in terms of BER accuracy and matching time. The performances are shown in Tab. 1. The matching time refers to the time passes to match a query image with a single image (pair-wise) in the dataset.

According to Tab. 1, the proposed method decreased the matching time dramatically. On the other hand, due to the decrement in the number of features, the BER accuracy showed a very small decrease which is less than 1%. We applied the proposed method in two more versions. We extracted the new features for two half parts and the entire shape without partitioning. The accuracy is naturally decreased along with the decrement in feature amount however, the computation time is also decreased. When we observe the BER (r_{BE} and computation time of the recent studies, we can observe that our method gives satisfactory results. In terms of BER, the proposed method with quadrants (4 pieces), halves (2 pieces), and single-piece achieved 89.6%, 87.9%, and 80.44% in image retrieval. The highest score belongs to AIR+LCDP+TSR [30] with 100% BER however, this study has two stages in which they eliminate unrelated images classed in the first one. Therefore they could achieve very high accuracy. The studies IDSC+SC+SCA [31] and IDSC+GMM-r [32] have higher BER than the proposed method however their computation time is much higher than our method. In literature, there is a large number of studies that used MPEG-7 dataset for benchmark however, we did not include the ones which are outdated or do not contain matching time in their results.

A similar test is conducted on the ETH-80 dataset with the leave-one-object-out method for performance

measure. In this dataset, recognition is used instead of retrieval. The recognition rate (RR) and computation time are given inTab. 2. The proposed method gives a similar outcome to the former dataset. The matching time decreased compared to the conventional method [21] while the RR is at the level of state-of-art.

There are more studies using ETH-80 dataset however some of them use the color as the cue or use different experimental methods rather than leave-one- object-out. We compared our results with the studies using the same method with us and shape contour as the cue. Besides, not all the studies inTab. 2 contain information related to their pair-wise matching time. Therefore, we represent those studies as NA in Tab. 2.

6 CONCLUSION

In this paper, an extended version of our previously published work is presented. The motivation is to obtain a state-of-art level recognition/retrieval accuracy with decreased computation time. Statistical features are extracted from the contour of shapes after the shape is divided into quadrants. Each shape is represented by a small size feature matrix. The proposed method was evaluated on MPEG-7 Part B and ETH-80 datasets according to the specific metrics of each dataset. The results show that the proposed method supplies a significant decrease in computation time and satisfying retrieval accuracy at the state-of-art level.

References

[1] A. Sezavar, H. Farsi, and S. Mohamadzadeh, "A Modified Grasshopper Optimization Algorithm Combined with CNN for Content Based Image Retrieval", *International Journal of Engineering*, vol. 32, no. 7, pp. 924–930, 2019.

[2] A. Feizi, "Convolutional Gating Network for Object Tracking", *International Journal of Engineering*, vol. 32, no. 7, pp. 931–939, 2019.

[3] M. Keyvanpour, R. Tavoli, and S. Mozaffari, "Document Image retrieval based on keyword spotting using relevance feedback", *International Journal of Engineering*, vol. 27, no. 1, pp. 7–14, 2014.

[4] A. F. Smeaton, E. O'Connor, and F. Regan, "Multimedia information retrieval and environmental monitoring: Shared perspectives on data fusion", *Ecological informatics*, vol. 23, pp. 118–125, 2014.

[5] K. H. Hwang, H. J. Lee, and D. J. Choi, "Medical Image Retrieval: Past and Present", *Health Informatics Research*, vol. 18, no. 1, 2012.

[6] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A Review of Content-Based Image Retrieval Systems in Medical Applications - Clinical Benefits and Future Directions", *International Journal of Medical Informatics*, vol. 73, no. 1, 2004.

[7] J. S. Hong, H. Y. Chen, and J. Hsiang, "A Digital Museum of Taiwanese Butterflies", *In Proceedings of the Fifth ACM Conference on Digital Libraries*, pp. 260–261, San Antonio, Texas, United States, ACM Press, 2000.

[8] B. Zhu, M. Ramsey, and H. Chen, "Creating a Large-Scale Content-Based Airphoto Image Digital Library", *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 163–167, 2000.

[9] R. Torres and A. Falco, "Content-Based Image Retrieval: Theory and Applications", *RITA*, vol. 13, pp. 161-185, 2006.

[10] H. Shao, Y. Wu, W. Cui, and J. Zhang, "Image retrieval based on MPEG-7 dominant colour descriptor", *Proceedings of the 9-th International Conference for Young Computer Scientists*, ICYCS, pp. 753–757, 2008.

[11] X. Duanmu, "Image retrieval using color moment invariant", *Proceedings of the Seventh International Conference on Information Technology: New Generations*, (ITNG), pp. 200–203, IEEE, Las Vegas, NV, USA, 2010.

[12] G. H. Liu, Z. Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on micro-structure descriptor", *Pattern Recognition*, vol. 44, No. 9, pp. 2123–2133, 2011.

[13] X. Y. Wang, Z. F. Chen, and J. J. Yun, "An effective method for color image retrieval based on texture", *Computer Standards & Interfaces*, vol. 34, no. 1, pp. 31–35, 2012.

[14] S. Fadaei, R. Amirfattahi, and M. R. Ahmadzadeh, "Local derivative radial patterns: a new texture descriptor for content-based image retrieval", *Signal Processing*, vol. 137, pp. 274–286, 2017.

[15] R. Sablatnig *et al*, "A novel image retrieval based on visual words integration of SIFT and SURF", *PLoS One*, vol. 11, no. 6, Article ID e0157428, 2016.

[16] D. Zhang and G. Lu, "Review of shape representation and description techniques", *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, 2004.

[17] J. M. Guo, H. Prasetyo, and J. H. Chen, "Content-based image retrieval using error diffusion block truncation coding features", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 466–481, 2015.

[18] H. Zhang, Z. Dong, and H. Shu, "Object recognition by a complete set of pseudo-Zernike moment invariants", *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, (ICASSP), pp. 930–933, IEEE, Dallas, TX, USA, 2010.

[19] N. Arica and F. T. Y. Vural, "BAS: a perceptual shape descriptor based on the beam angle statistics", *Pattern Recognition Letters*, vol. 24, pp. 9–10, 2003.

[20] M. E. Yıdırım, Ö. F. Ince, Y. B. Salman, E. Sadıç and J. Park, "Recognition of Single-Land Countries on Outline Images by Using BAS Feature", *18th International Conference on Control, Automation and Systems*, pp. 656–660, 2018.

[21] M. E. Yildirim, O. F. Ince, B. S. Yucel, and I. F. Ince, "Shape retrieval using angle-wise contour variance", *Journal of Electrical Engineering*, vol. 72, no. 2, pp. 99-105, 2021.

[22] X. Wang, B. Feng, X. Bai, W. Liu, and L. Latecki, "Bag of contour fragments for robust shape classification", *Pattern Recognition*, vol. 47, no. 6, pp. 2116-2125, doi.org/10.1016/j.patcog.2013.12.008, 2014.

[23] C. Lin, C. M. Pun, and C. M. Vong, "Efficient shape classification using region descriptors", *Multimedia Tools and Applications*, vol. 76, pp. 83–102, 2017.

[24] F. Ni and B. Wang, "Integral contour angle: An invariant shape descriptor for classification and retrieval of leaf images", *Proceedings of the 25th IEEE International Conference on Image Processing*, (ICIP), pp. 1223–1227, Athens, Greece, 2018.

[25] A. G. Maidana, L. A. Horacio, C. Schaerer, and V. V. Waldemar, "Contour-point signature shape descriptor for point correspondence", *International Journal of Image and Graphics*, vol. 18, no. 2, 1850007, 2018.

[26] P. Govindaraj and M. Sudhakar, "A new 2D shape retrieval scheme based on phase congruency and histogram of oriented gradients", *SIViP*, vol. 13, pp. 771–778, 2019.

[27] L. J. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, IEEE, pp. 424–429, 2000.

[28] N. G. de Sá, L. P. Valem, and D. C. G. Pedronette, "A Multi-level Rank Correlation Measure for Image Retrieval", *Proceedings of International Conference on Computer Vision Theory and Applications* VISAPP , vol. 5, pp. 370–378, 2021.

[29] B. Leibe and B. Schiele, "Analyzing Appearance and Contour Based Methods for Object Categorization", *CVPR*, vol. 2, pp. 409–415, 2003.

[30] X. Pan, S. Chachada, and C.-C. J. Kuo, "A two-stage shape retrieval (TSR) method with global and local features", *Journal of Visual Communication and Image Representation*, vol. 38, pp. 753–762, 2016.

[31] S. Bai and X. Bai, "Sparse contextual activation for efficient visual reranking", *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1056–1069, 2016.

[32] P. Kontschieder, M. Donoser, and H. Bischof, "Beyond pairwise shape similarity analysis", *Asian conference on computer vision*, pp. 655–666, 2009.

[33] X. Bai, C. Rao, and X. Wang, "Shape vocabulary: a robust and efficient shape representation for shape matching", *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3935–3949, 2014.

[34] Y. Zheng *et al*, "Fourier Transform to Group Feature on Generated Coarser Contours for Fast 2D Shape Matching", *IEEE Access*, vol. 8, pp. 90141–90152, 2020.

[35] P. F. Felzenszwalb and J. D. Schwartz, "Hierarchical matching of deformable shapes", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (CVPR) pp. 1–8, 2007.

[36] B. Wang and Y. Gao, "Hierarchical string cuts: A translation, rotation, scale, and mirror invariant descriptor for fast shape retrieval", *IEEE Transactions of Image Processing*, vol. 23, no. 9, pp. 4101–4111, 2014.

[37] N. Alajlan, I. E. Rube, M. S. Kamel, and G. Freeman, "Shape retrieval using triangle-area representation and dynamic space-warping", *Pattern Recognition*, vol. 40, no. 7, pp. 1911–1920, 2007.

[38] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[39] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, 2007.

[40] N. Kaothanthong, J. Chun, and T. Tokuyama, "Distance interior ratio: A new shape signature for 2D shape retrieval", *Pattern Recognition Letters*, vol. 78, pp. 14–21, 2016.

[41] R. Hu, W. Jia, H. Ling, and D. Huang, "Multiscale distance matrix for fast plant leaf recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pp. 4667-4672, 2012.

[42] B. Ramesh and C. Xiang, "Unseen object categorization using multiple visual cues", *Neurocomputing*, vol. 230, pp. 88-99, 2017.

[43] M. R. Daliri and V. Torre, "Shape recognition based on kernel-edit distance", *Computer Vision and Image Understanding*, vol. 114, pp. 1097–1103, 2010.

[44] M. Daliri and V. Torre, "Robust symbolic representation for shape Recognitition and retrieval", *Pattern Recognition*, vol. 41, no. 5, pp. 1782–1798, 2008.

[45] J. Wang, X. Bai, X. You, W. Liu, and L. J. Latecki, "Shape matching and classification using height functions", *Pattern Recognition Letters*, vol. 33, pp. 134–143, 2012.

Received 2 June 2022