Baghdad Science Journal

Volume 22 | Issue 1 Article 30

2025

A Novel Approach for Shape Pattern Recognition Based on Boundary Features Generated by Line Simplification Algorithm

Ali Adel Saeid
Computer Science Department, University of Technology, Baghdad, Iraq, ali.a.saeid@uotechnology.edu.iq

Raheem Ogla
Computer Science Department, University of Technology, Baghdad, Iraq,
Raheem.a.ogla@uotechnology.edu.iq

Shaimaa H. Shaker
Computer Science Department, University of Technology, Baghdad, Iraq,
shaimaa.h.shaker@uotechnology.edu.iq

How to Cite this Article

Saeid, Ali Adel; Ogla, Raheem; and Shaker, Shaimaa H. (2025) "A Novel Approach for Shape Pattern Recognition Based on Boundary Features Generated by Line Simplification Algorithm," *Baghdad Science Journal*: Vol. 22: Iss. 1, Article 30.
DOI: 10.21123/bsj.2024.9517
Available at: https://bsj.researchcommons.org/home/vol22/iss1/30

BAGHDAD SCIENCE JOURNAL 2025;22(1):349–360

RESEARCH ARTICLE

A Novel Approach for Shape Pattern Recognition Based on Boundary Features Generated by Line Simplification Algorithm

Ali Adel Saeid *, Raheem Ogla, Shaimaa H. Shaker

Computer Science Department, University of Technology, Baghdad, Iraq

ABSTRACT

Shape recognition is an essential task in machine vision applications. Many techniques have been adopted for shape recognition all of them distributed according to two directions of shape features belonging to boundary or region, in many categories of applications in addition to recognition like shape simplification, restoration classification and retrieving. This paper presents a proposed testing algorithm for shape recognition according to boundary or global base features where the approximated contour of the shape is derived by the Douglas Peucker line simplification approach. The initial diversity value starts with fixit value according to an undetermined number of iterations until reaching to target set of approximated points for testing, where in each iteration the diversity value increases gradually with the fixit interval. This algorithm evaluates the state, "Is there an optimal number of approximated points as features vector can be adopted that satisfy the maximum recognition rate." Before learning and testing stat the features vector derived after applied a distance matrix among the approximated set. The recognition process experiments on the MPEG-7 dataset when classified using Support Vector Machine (SVM) through different kernel functions. However, the experiment results show the proposed testing recognition achieved to high rate about 0.961 according to a specific number of approximated datasets when adopting a radial basis function (RBF) as kernel function based on Matlab testing environment.

Keywords: Douglas Peucker, Line simplification, Pattern recognition, Shape recognition, SVM

Introduction

Shape recognition is an essential problem in image classification where each shape represents an individual class.[1] Recognizing shapes occupy an important field in computer scientific research like image understanding, pattern recognition, robotic vision system, and document image analysis.[2–4] Shape recognition problems can be classified into two kinds (online and offline). Online shape recognition is related to the recognition process while the shape is being generated on the contrary offline recognition deals with scanned or printed shapes.[5,6] Therefore, this process must be satisfying three sequential states starting from (data preprocessing, features extraction and classification).[6] Number of feature extraction techniques area developed for the shape recognition process, like (color histogram, Eigenvector Approaches, SIFT-Scale Invariant Feature Transform, Histograms of Edge Directions (HED), and Shape Descriptor). However, this research is concerned with the last technique (Shape Descriptor). Generally, this technique is classified according into two efficient approaches, (region-based, contour-based).[7–9] and each of them classified into two approaches, global approach and structural approach. With the contour-based all the boundary pixels are taken into consideration for extra features, and not subject to any division process that agrees with the global approach on the opposite state with the structural approach all boundary pixels subject to the division process which converts the boundary points to segments or

Received 18 September 2023; revised 25 February 2024; accepted 27 February 2024.
Available online 1 January 2025

* Corresponding author.
E-mail addresses: ali.a.saeid@uotechnology.edu.iq (A. A. Saeid), Raheem.a.ogla@uotechnology.edu.iq (R. Ogla), Shaimaa.h.shaker@uotechnology.edu.iq (S. H. Shaker).

https://doi.org/10.21123/bsj.2024.9517

primitives.[10,11] Most of previous studies related on shape matching and recognition were based on boundary metrics features like (contour based, contour flexibility and shape context), the sampling process depends on the initial step after extracting the boundary of features and sample with equal space, and the complexity of local features. Previous studies relied on (equal spacing) did not adopt the optimal method for sampling boundary points to select the most suitable number of approximated points that accurately represent the original boundary features.[11–15] However, the number of points has a notable impact on the factor: the quality of recognition. Therefore, this paper proposed a novel approach for deriving a shape descriptor according to (contour-based) through by extra subset of boundary points by line simplification technique according to Douglas Peucker algorithm for derive a features vector to evaluate the given assumption, "one needs to investigate whether there is an optimal number of approximated points within a subset of all boundary points that satisfies a maximum recognition rate". Where the support vector machine is adopted for nonlinear classification tasks depending on the kernel function mathematical approach. This paper is organized as follows: The next section presents a related work. Section three introduces the theoretical background. The proposed recognition testing algorithm was represented in section four. Section five reviews the experimental results and the conclusion illustrated in the last section.

Related work

This section presents briefly related works for shape recognition as follows: Yang *et al.*[16] suggested a novel approach to shape recognition according to combining processes after dividing the shape's boundary contour points into two levels (contour fragments and sampling points) to represent and describe the global and detailed information of the shape sequentially. Each level will be subject to an encoded process by the Fisher vector approach and combined them to perform a shape recognition process through linear (SVM) model. The accuracy results from range values up 92.70% to 99.26 according to different datasets. Yan *et al.*[17] presented a study of grain sand's shape recognition, related to the comparison state between Fourier descriptors and shape descriptors (circularity, elongation, roughness, and convexity) to determine the optimal discriminant one of them (according to the relationship between manufactured and natural sand). Firstly, the shape descriptors metrics is applied and statistical features will be computed for each(grains of sand and in another comparison side the Fourier descriptors is computed to reconstruct the boundary shape of grains of sand. The Andrews plots are adopted for each side as a discriminant plotted function. Therefore, the high quality of Fourier descriptors is dependent. Wen-Yen[18] proposed a shape recognition approach where the boundary edge of the object is extracted to determine the dominant set of approximated points. These constative points are a global recognition feature, which is categorized as a nonlinear segment or liner segment for constructing compactness approximate- polygons to derive discriminant features, where the cyclic string matching recognition method is adopted. The maximum recognition rate is starting from 75% to 96 depending normalization methods on. Li Z *et al.*[19] presented shape recognition method that depends on a richness of local contour features (contour orientation, contour pixel, and contour distance) where initially the local contour features are obtained by the minimum bounding rectangle method. González *et al.*[20] proposed a shape descriptor based on curvature features for boundary points like (radius, center and tangent) and quantized the features vector by k-means algorithm according to minimizing the centroid distance between classes the maximum recognition rate obtained by this research equal 86.3% where MPEG-7 dataset is adopted in testing. Mirehi *et al.*[21] introduced a study of constructing a GNG graph for object shape recognition that captures the geometrical and topological properties features of the object and relational features between internal and boundary Vertices. Dynamic programming (DP) is adopted to determine the optimal matching. The experiment results showed the high ability of this method for recognition against noises depending on topological graph features. Paramarthalingam and Thankanadar.[22] proposed one dimension shape features vector disrupter based on boundary contour points to derive six disrupters. The experiment results showed the object area normalization (OAN) descriptor has a better performance compared with other disrupters according to shape marching and image retrieving where MPEG-7 dataset is adopted in testing. Aswathi and Philumon[23] suggested an object shape recognition method based on combining the bag–of–word model with multi-scale curvature descriptor where the classification stage supports the vector machine model. The recognition accuracy value is equal to 80%. Zheng *et al.*[24] proposed a novel framework of global boundary furious disrupters, (multiscale –Fourier-descriptor using -group feature) MSFDGF. The experiments were applied on four standard databases and the result shows a high accuracy result is 87.76%. with MSFDGF-SH descriptor compared with another

Fourier descriptor. Ahmed and Aradhya.[25] proposed a technique for recognizing the shape of object based on extracting eight neighborhood patterns according to state position for each boundary counter pixels with assigning a unique label. The matching process between shapes is based on the minimum cost of hit value. The proposed technique is experimented on the standard datasets where recognition accuracy is 98%. Zheng *et al.*[26] presented a global shape descriptor called (IDSC-wFW) a weighted -Fourier and wavelet like descriptor according to fast matching properties, based on inner-distance shape–context. The initial state starts by changing the shape histogram (IDSC descriptors) belonging to the point domain to the histogram based on the field (as one-dimensional signal model) and transforming the last histogram to a new one-dimension signal transform based on the Fourier transform. Lastly, the final two transforms are combined to create a new descriptor. The experiment result showed that the proposed descriptor technique is efficient for use in online retrieval or large databases. Zheng *et al.*[27] proposed a new frequency-domain Fourier descriptor (FMSCCD) based on (centroid-contour-distance, Fourier transform and multiscale description). The experimental results showed that it's an efficient discriminable and simple descriptor model with low computational cost. Rababaah and Rabaa.[28] presented a method for shape recognition and characterization based on (polar shape-signature and template-matching technique). The purpose of this proposed method is to support the application of smart vehicles and robots. The proposed algorithm achieved robust accuracy equal to 95.83% and *95.45%* when testing the testing shapes and hand-drawing shape respectively.

Materials and methods

Line simplification

Line simplification defined as the technical process of reducing redundant points of a complex line while maintaining the vital points of essential shape characteristics.[29,30] Today, this technique has many applications like shape analysis, cartography, topological-consistency assessment, and data quality.[31] Line simplification algorithms can be classified as follow according to the selection approach for boundary points.[32]

Independent-point approach

With this algorithm approach the process of removing redundant points depends on predefined condition without taking account of the relationship among the boundary points, for example (n^{th} point algorithm).

Local-processing routines

The algorithms with category take into the consideration the relationship between the consecutive pints according to distance or perpendicular distance. The algorithms with this category include ones like (Reumann Witkam Routine, Triangular-Routine, and Lang-Routine).

Global routine

The global processing routines differ from the local approach through Reumann Witkamtial data points are subject to entirety processing. This research presents the **Douglas Peucker** algorithm as a local line simplification technique.

Douglas-Peucker algorithm

Douglas and Peucker proposed a line simplification algorithm which used in many applications like path planning problems, geographic applications and data compression.[33–35] The fundamental object of this algorithm is to re-represent the polyline of the curve, to a similar liner curve according to piecewise segments, which leads reducing the number of the original curve's points. Suppose the original trajectory of the curve is defined as, $K = (k_1, k_2, \ldots k_i, \ldots, k_n)$, this trajectory can be substituted by line segments $\overline{k1k2}, \overline{k2k3}, \ldots, \overline{ki-1ki}, \ldots, \overline{kn-1kn}$. Therefore, the original trajectory curve is reconstructing by fewer segments based on selected points or vertices from the original points[36] Fig. 1. Initially, the original data handled by the algorithm are trajectory points of the curve and threshold distance value or (diversity value or parameter) ($\epsilon > 0$) was predefined.[37]

The initial step of this algorithm starts by keeping the endpoints (p_1 *and* p_N). The algorithm determines the optimal point p_k from (p_2 *to* P_{N-1}), where at this point the maximum error (perpendicular distance) β_k from the connected line between the endpoints (p_1, p_N) is greater than or equal to the predefined tolerance value (ϵ),[38] the point is marked "keep". This operation recursively applied[38] by splitting the curve at pk as two new curve segments C_1 $(p_1 \ldots p_k)$, C_2 $(p_k \ldots p_N)$ as shown in Fig. 1. Finally, the recursion process terminates if the maximum error βk is less than (ϵ) or the polyline decreases to a segmented line with only two vertices. Finally, the approximated line consists of all points marked as "keep."[39] Following pseudocode algorithm steps of Douglas-Peucker Algorithm.[40]

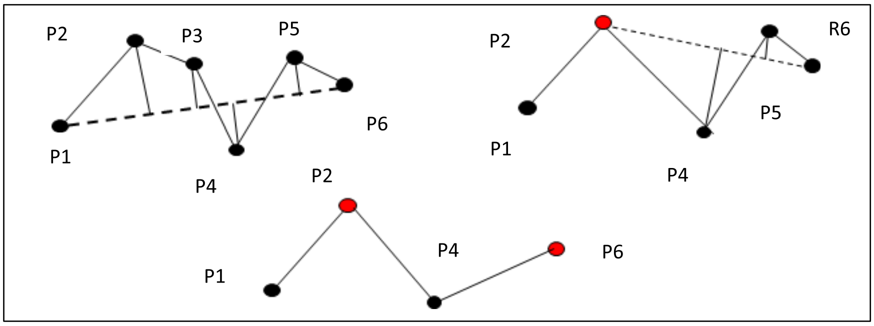


Fig. 1. Sequential steps after applied Douglas-Peucker algorithm.

Algorithm 1. Peucker algorithm

Input: trajectory of boundary points. Output: Approximated datapoint.
Step 1: Let a curve C composed of N Vertices set defined as $V = \{v_1, v_2, \ldots, v_n\}$, and diversity value $\epsilon = 0.001$; Step 2: connect the first V_1 and last Vertices V_n to obtain a straight line $C_{v1,vn}$, and calculate the minimum distances between a string line $C_{v1,vn}$ and the remaining points (Vertices), $\{V_2, \ldots V_{n-1})$ where the distance set defined as $S = \{S_2, \ldots, S_m, \ldots S_{n-1}\}$. Step 3: determine the maximum distance S_{max} from distance set S, $S_{max} = S_m$ where S_m is the distance between straight line and Vertis V_k. Step 4: if ($S_{max} < \epsilon$) then Vertices set, $\{V_2, \ldots V_{n-1})$ are rejected and the given curve is Compressed to straight line $C_{v1,vn.}$ and go to end. Else the original vertices set $V = \{v_1, v_2, \ldots, v_n\}$ is divided into two subsets V_t and Vs, where $V = V_t + Vs$, $V_{t=} = \{V_1, V_2, \ldots, V_m\}$, $V_{S=} = \{V_m, V_{K+1}, \ldots, V_n\}$. Step 5: For each subset V_t and Vs, apply the previous steps recursively until any distance value is less than the diversity value ϵ. Step 6: end.

Support vector machine

The support vector machine (SVM) was invented in 1963, It is a supervised learning classifier algorithm[41,42] used for the classification and recognition process. The target of this learning algorithm is specified the optimal hyperplane with maximum- margin in N-dimensional space for separating data,[39] mathematically let training dataset is defined as $S = \{(v_1, k_1), (v_2, k_2), \ldots (v_n, k_n)\}$, $vi \in \{1,-1\}$, and the defended hyperplane $z = w^T k + b = 0$ which adopted for separate samples according to different classes after selective a descriptive parameters[43,44] where w is the orientation parameter of a hyperplane, k is a point lying on the hyperplane and b is the bias of the distance of hyperplane starting from the origin.[44] Therefore, the optimal hyperplane can satisfy the separation process as follow equations:[44]

$$Z = \begin{cases} w^t\ k + b \geq 0, yi = +1 \\ w^t\ k + b \leq 0, yi = -1 \end{cases} \quad (1)$$

SVM algorithm seeks to find the maximum distance between the hyperplane and the learning class's points which are called margin distance.[45] However, with a simple linear separable state the SVM algorithm finds the liner separating hyperplane which maximizing classifier margin according to two classes Fig. 2(a).[43] In contrast in a nonlinear state when the classes cannot be separated in linear state the classification process is applied in new dimension space after deriving a suitable hyperplane classification model where the data is subject to nonlinear transform (Fig. 2(b,c)).[43] The optimization problem is derived according to maximize the margin vertical distance which defined as.[43,45]

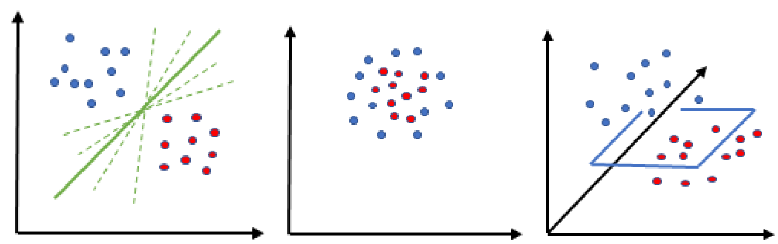
$$\text{Max}_{w,b} = \frac{2}{||w||^2} \quad (2)$$

That mean the maximizing process is based on minimizing ||w|| therefore.

$$\text{Min}_{w,b} = \frac{1}{2}\ ||w||^2 \quad (3)$$

Now for simplify the previous complex problem depend on Lagrange function lead to the following modified equation.[41]

$$L_{(w,b,a)} = \frac{1}{2}\ ||w||^2 - \sum_{i=1}^{m} a_i y_i \left(w^t\ k + b\right) + \sum_{i=1}^{m} a_i \quad (4)$$



(a) The liner separable problem (b) Points distributed according to nonlinear state (c) The nonlinear separable state

Fig. 2. Support vector machine models.

Table 1. Kernel functions.

Kernel functions	form
Liner	$P(Z_i, Z_j) = Z_i^T.Z_j$
Polynomial	$P(Z_i, Z_j) = (Z_i^T.Z_j + 1)^d$
Gaussian	$P(Z_i, Z_j) = \exp(-\frac{\|\|(Zi-Zj\|\|2}{\sigma 2})$

With taking the partial derivative of b and w in the Eq. (4) and after number deriving state by KKT (Karush, Kuhn, Tucker) the next minimizing equation is as follows.[44]

$$\text{Min}\,\frac{1}{2}\;\|w\|^2 + c\sum_{i=1}^{n}\mathcal{E}_i\,\forall_i\;yi\left(w.x_i + b\right) > 1 - \mathcal{E}_i \quad (5)$$

Where and $\mathcal{E}_i$ is a slack parameter that is dependent on selecting a hyperplane with less cost and error, c is the regularization parameter.[43,45] Some samples cannot be classified according to suitable hyperplane when they project in higher dimension space from the original sample space, therefore for solve this problem these samples are mapped to new higher dimension space according to kernel functions (kernel trick) as follows (Table 1) and dual problem will be as in Eq. (3).[43–47]

$$\max_{\pounds}\sum_{i=1}^{\pounds}\pounds i - \tfrac{1}{2}\sum_{i,j=1}^{\pounds}\pounds i\pounds jyiyj\,P\left(xi, xj\right)\;\forall i, \qquad (6)$$
$$0 \leq \pounds i \leq c\sum_{i=1}^{n}\pounds jyi = 0$$

where $\pounds$i, Lagrange –multipliers, is a kernel function.

To choose the optimal one of them (kernel function), the user must apply the testing process to determine the suitable one of them.

The propose approach

This section includes the proposed recognition algorithm that initially depends on extract the approximated set of points from the contour boundary of the shape according to Douglas-Peucker algorithm. Firstly, starting with an initial value of diversity (ϵ) and increasing it with each approximation iteration by the DP algorithm until reaching to specific predefined number of approximated points, that agreed with the required size of feature vector points. During this processing, sometimes the numbers of achieved points (approximated points) is closed to required numbers. Therefore, the approximate set of points is submitted to regulating evaluation process for reaching to required number, through by sequential computing process to determine the radiates value (R_i) for each approximated point. This value is computed according to the set of the created circle's equations of the approximated point and its neighbors located on the circumference of the virtual circle Fig. 3. Now the approximate set of points that exceed the required number with lower radios from a low curvature is removed to ensure that the final number of points matches the required number of features vectored size. However, now the approximated set of points is subject to the computed state by applauding distance matrix for generating features vector according to non-duplicated values selected from a symmetric previous matrix to classification step.

Results and discussion

This section presents the evaluation stage for the recognition proposed algorithm mentioned before, according to the standard dataset MPEG-7 by MATLAB_R2017b. Initially, the first evaluation deals with determining the assumption (is there an optimal number of approximated boundary set of points adopted for the shape recognition process with considering a specific image size, 200 × 200), among different numbers of approximated sets of points. The adopted classifier is an SVM and the

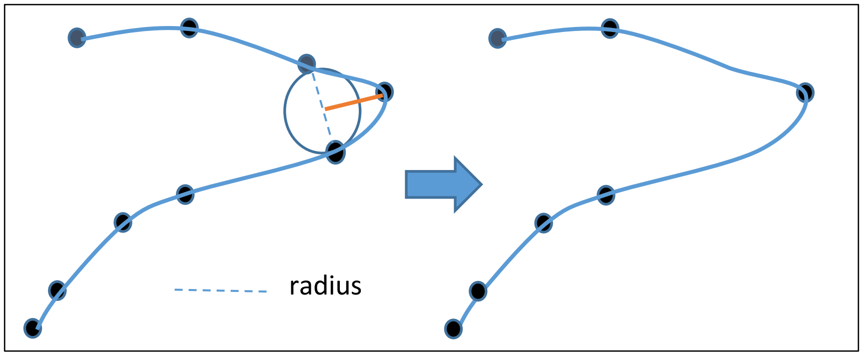


Fig. 3. Regulating evaluation process.

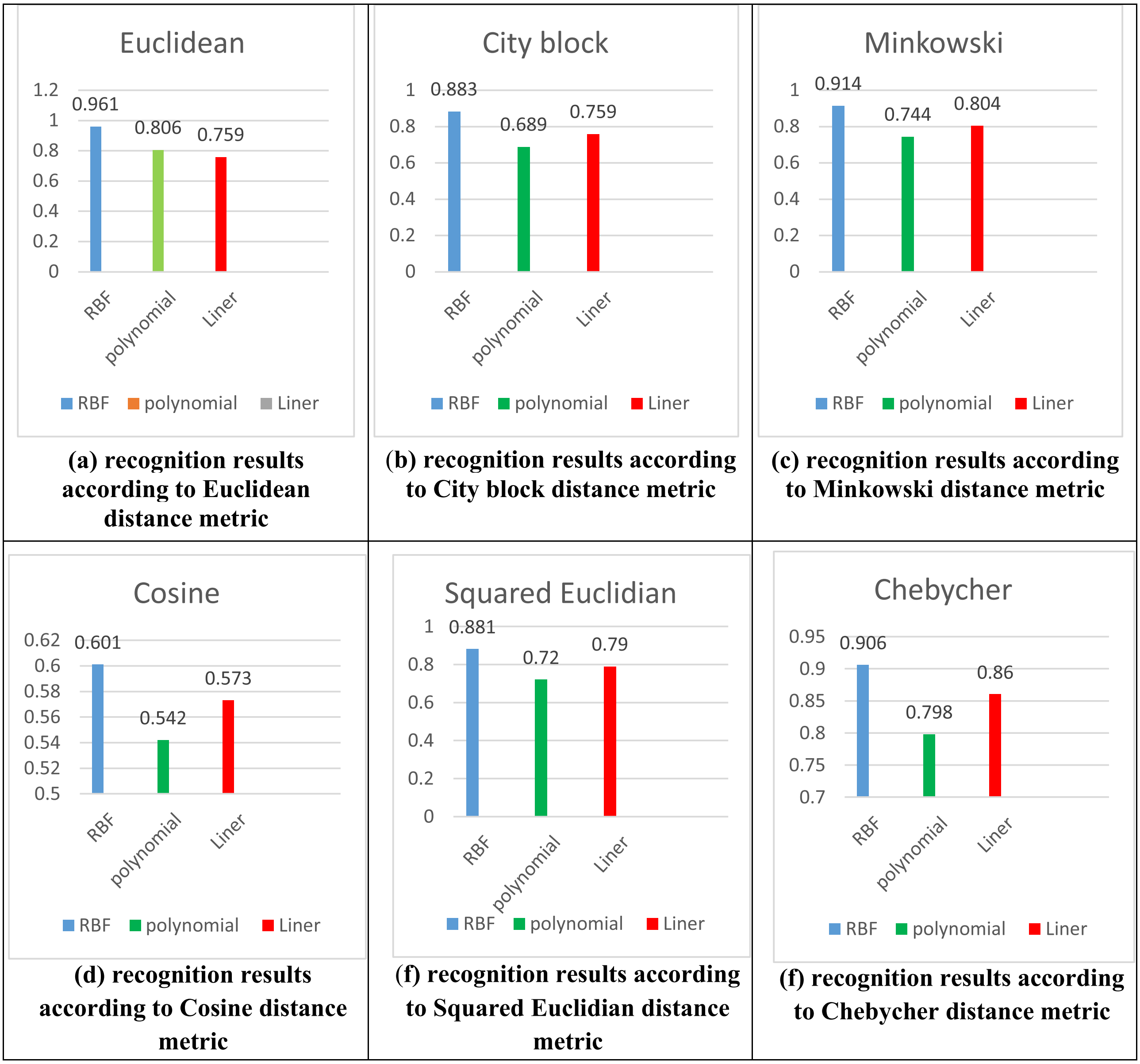
Algorithm 2. The recognition proposed algorithm

Input: tested image.
Output: image shape class label.

Step 1: let the diversity value $\alpha = 0$, the increasing value, $\epsilon = 0.001$, features vector requirements points size β. Number of approximated points np = 0;
Step 2: Read the tested image I.
Step 3: Extra the shape boundary SB.
Step 4: repeat
Step 4.1: $\alpha = \alpha + 0.01$;
Step 4.2: np = Douglas-Peucker(α, SB);
Step 4.3: until (np $\geq \beta$).
step 5: If (np $= \beta$) go to step 6
step 5.1: else
step 5.2 T $= \beta$-np, u = 1
step 5.3 For each triple sequential point **np** calculate the virtual diameter circle.
step 5.4 Reject the approximated points (T) that meet the lowest diameter values
End if
Step 6: Calculate distance matrix (dissimilarity matrix) for an approximated point β.
Step 7: Calculate the cumulative summation distances for each point from the previous step and determine the selected one (sp) that meets the largest summation depended to solved rotation invariant state
Step 8: Apply the normalization process to the solved scaling invariant state.
Step 9: Generate the feature vector based on values in the upper trigonal distance matrix.
Step 10: Applied the SVM classifier to determine the class label.

kernel functions as a (radial base function (RBF), polynomial and Liner), with different ranks of polynomial's degree (cubic, Quadratic and Quartic). Initially, the Euclidian distance metrics were adopted by dissimilarity matrix for the recognition process, where ***n*** equals the number of approximated points and the size of the features vector equal to ((n*n) − n)/2, as illustrated in the following results (Table 2). However, because a dissimilarity matrix is symmetric and each diagonal element is equal to zero. The features vector will compose of all element values in the upper triangular dissimilarity matrix which (representing all distance values between each approximated point to all other points). Therefore, to evaluate our previous assumption we test the proposed algorithm according to different sizes of an approximate set of points derived by the (Douglas-Peucker) Algorithm starting from the initial size 10 to 30 with different types of kernels function as below.

As seen in the previous table, the maximum recognition rates are (0.961, 0.657, 0.793, 0.728, 0.620) satisfied when the approximated data points are 20 and 21 sequentially opposite each kernel function. However, the next step to evaluate the effect type of distance metrics on the recognition rate which is adopted by distance matrix (dissimilarity matrix). Therefore, for the next evaluating step adopted the set of approximated points, where (n = 20) is based on the RBF kernel function. However, for the second consideration after determining the optimal approximation points (20) in the previous state now we evaluate the rate of recognition process according to applied different distance metrics for creation a features vector by dissimilarity



(a) recognition results according to Euclidean distance metric

(b) recognition results according to City block distance metric

(c) recognition results according to Minkowski distance metric

(d) recognition results according to Cosine distance metric

(f) recognition results according to Squared Euclidean distance metric

(f) recognition results according to Chebycher distance metric

Fig. 4. The histograms of shape recognition results.

matrix. Fig. 4 illustrates the histograms recognition levels classified in three categories of SVM kernel functions.

From above Fig. 4, the maximum recognition accuracy values achieved (0.961, 0.914, 0.906) are satisfied when the kernel function is RBF meets the (Euclidian, Minkowcki and Chebycher) distance metric, sequentially according to each histogram. Generally, the observer can notice the higher accuracy recognition value with RBF kernel function and the liner kernel function are more recognition accurate than polynomial kernel function. Therefore, we can adopt the kernel classification function with the Euclidian distance metric according to it's the higher accuracy. However, bellow the observer can see the sample of recognition. shapes matching results (Fig. 5) according to leering patterns subset from original learning dataset, Fig. 6 where the white background figures represent the boundary approximated sets of points opposite to recognitions results by the proposed algorithm that matching appropriate suitable pattern in Fig. 5.

Fig. 7 illustrates the results of the proposed algorithm by using distorted shapes. Where is the

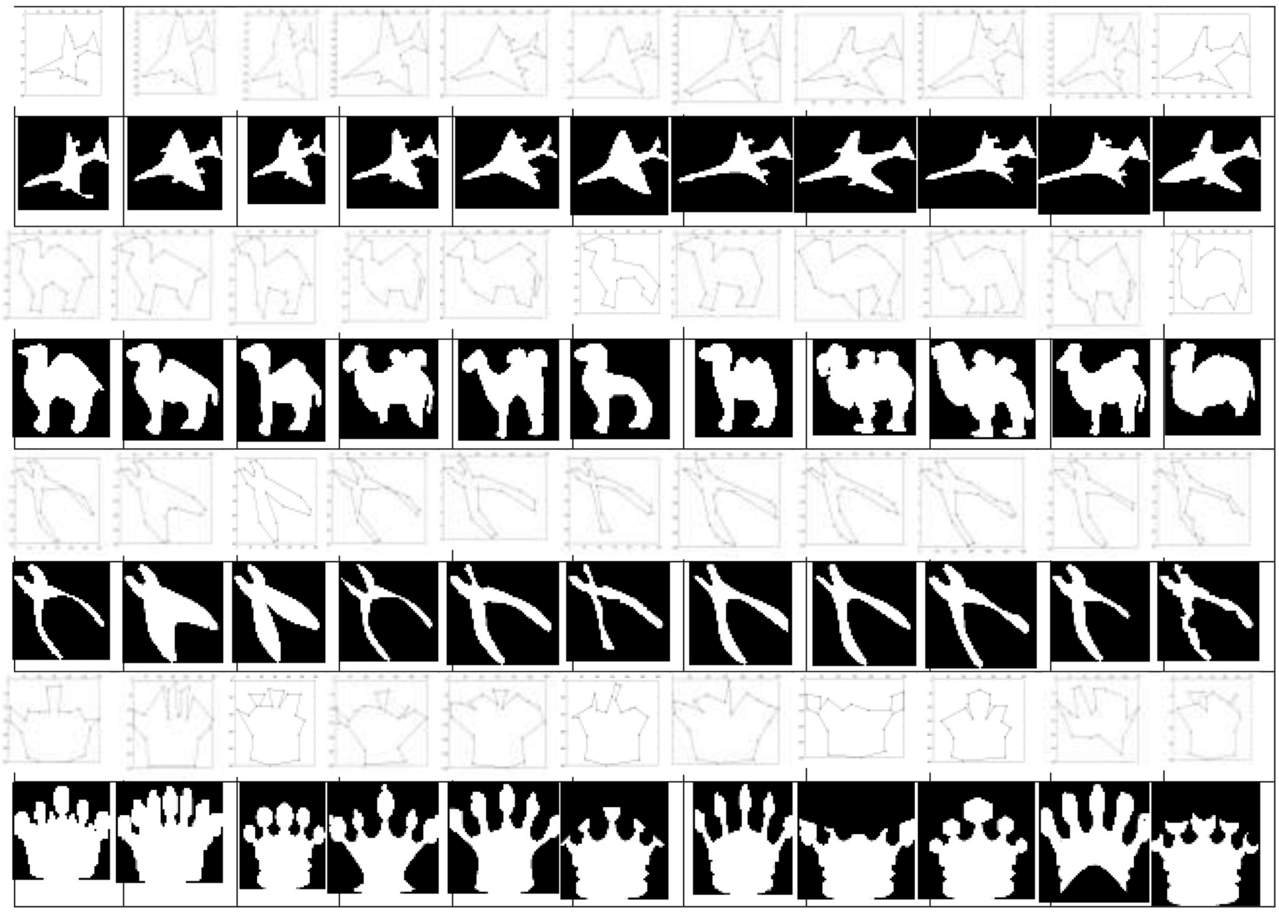


Fig. 5. Shows recognition testing patterns shapes opposite their approximated points are recognized by proposed algorithms according to next learning pattern where the white background figures represent the boundary approximated points opposite to recognitions results by the proposed algorithm.

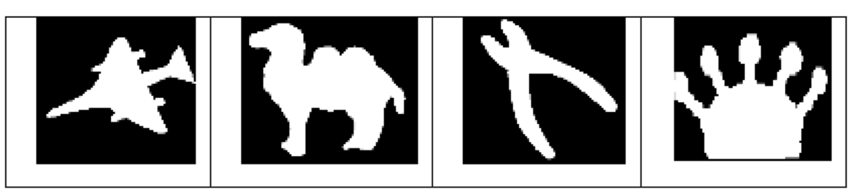


Fig. 6. Sample of learning shape patterns in MPEG-7 dataset.

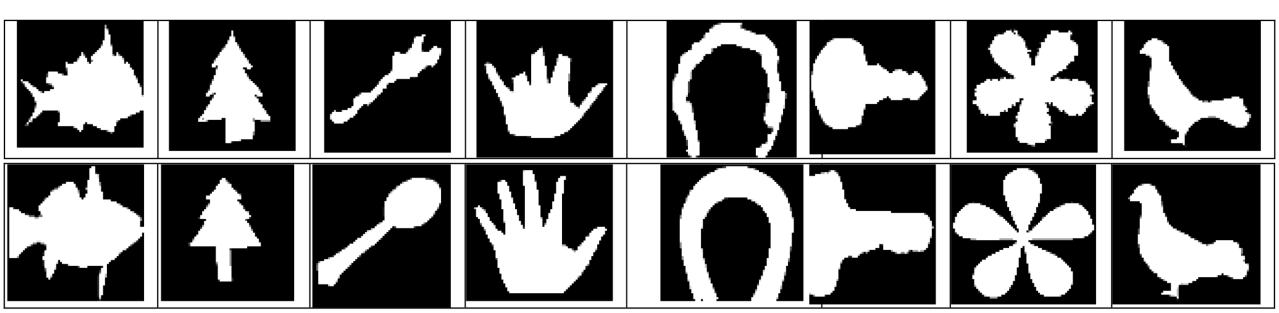


Fig. 7. The results of the proposed algorithm by using distorted patterns.

Table 2. Recognition accuracy results according to different sizes of approximated set of points.

Kernel function n (number of points)	Accuracy recognition values by **RBF** kernel function	Accuracy recognition values by **polynomial** kernel function	Accuracy recognition values by **liner** kernel function	Accuracy recognition values by **Quadratic** kernel function	Accuracy recognition values by **Quartic** kernel function
10	0.498	0.345	0.338	0.355	0.352
11	0.535	0.410	0.420	0.416	0.400
12	0.572	0.471	0.444	0.471	0.454
13	0.589	0.447	0.477	0.467	0.440
14	0.698	0.525	0.586	0.569	0.505
15	0.650	0.467	0.508	0.515	0.464
16	0.645	0.491	0.508	0.525	0.474
17	0.711	0.525	0.522	0.535	0.494
18	0.725	0.562	0.640	0.586	0.508
19	0.840	0.637	0.722	0.677	0.576
20	**0.961**	0,600	**0.793**	**0.728**	0.555
21	0.840	**0.657**	0.766	0.722	**0.620**
22	0.840	0.644	0.725	0.711	0.603
23	0.800	0.633	0.735	0.694	0.586
24	0.732	0.511	0.566	0.535	0.494
25	0.711	0.57	0.583	0.583	0.528
26	0.779	0.583	0.640	0.644	0.528
27	0.766	0.640	0.644	0.647	0.596
28	0.772	0.647	0.654	0.688	0.613
29	0.755	0.630	0.650	0.677	0.596
30	0.769	0.630	0.647	0.647	0.596

first row of figure representing the distorted patterns and after using the proposed algorithm, the output present in the second row of figure for improve the proposed algorithm.

Conclusion

This paper proposed a boundary or global-based descriptor algorithm based on line simplification approach according to selecting an optimal approximated set of points by the Douglas-Peucker algorithm. However, through the experiment side, the maximum recognition rate was achieved equal to 0.961 when the approximated dataset was equal to 20 where the support vector machine nonlinear classifier model was adopted with RBF a kernel model and distance metrics are (Euclidian, Minkowcki and Chebycher) was adopted by dissimilarity matrix. In future work, we will work on reducing the time complicity by optimizing the efficiency of algorithm for selection the optimal diversity value, and then applying the future work in real field like Hand Gesture Recognition.

Authors' declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for republication, which is attached to the manuscript.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at University of Technology.

Authors' contribution statement

A. A. S. performed the design, acquisition of data, analysis, interpretation, and drafting the MS. R O. A. did the revision and proofreading. S. H. S. did the revision and proofreading.

References

1. Wang X, Ding W, Liu H, Huang X. Shape recognition through multi-level fusion of features and classifiers. Granul Comput. 2019;5:437–448. https://doi.org/10.1007/s41066-019-00164-8.
2. Yang C, Yu Q. Invariant multiscale triangle feature for shape recognition. Appl Math Comput. 2021;403(6). https://doi.org/10.1016/j.amc.2021.126096.
3. Cong V, Hanh L, Phuong L, Duy D. Design and development of robot arm system for classification and sorting using machine vision. FME Trans. 2022;50:181–192. http://dx.doi.org/10.5937/fme2201181C.
4. Yadav AS, Kumar S, Karetla GR, Cotrina-Aliaga JC, Arias-Gonzáles JL, Kumar V, *et al.* Feature extraction using probabilistic neural network and BTFSC-net model with deep learning for brain tumor classification. J Imaging. 2022;9(1):1–22. https://doi.org/10.3390/jimaging9010010.

5. Mannan A, Babri H, Saeed M. Offline shape recognition using flexible DCT grid. Sci Iran. 2012;19(6):1722–1730. https://doi.org/10.1016/j.scient.2012.10.0.
6. Patel M, Tandel P. A Survey on Feature Extraction Techniques for Shape based Object Recognition. Int J Comput Appl. 2016;137(6):16–20. https://doi.org/10.5120/ijca2016908782.
7. Ahmed M, Aradhya M. 2D Shape Recognition and Retrieval Using Shape Contour Based on the 8-Neighborhood Patterns Matching Technique. IGI Global-(JACHI). 2019;10:49–61. http://dx.doi.org/10.4018/IJSE.2019070104.
8. Yan X, Yang M. A Comparative Study of Various Deep Learning Approaches to Shape Encoding of Planar Geospatial Objects. ISPRS Int J Geoinf. 2022;11:527. https://doi.org/10.3390/ijgi11100527.
9. Basavanna M, Prem Singh M, Prakash Raje Urs M, Chandraiah T. Recognition of Geometrical Shapes Using Inclination and Statistical Features. IJNIET. 2020;12:21–26.
10. Kaur R, Devendran V. Content Based Image Retrieval. IJITEE. 2020;9:222–228. https://doi.org/10.1080/23311916.2021.1927469.
11. YANG C, Fang L, Wei H. Learning Contour-based Mid-level Representation for Shape Classification. IEEE Access. 2020:157587–157601. https://doi.org/10.1109/ACCESS.2020.3019800.
12. Mori G, Belongie S, Malik J. Efficient shape Matching Using shape Contexts. IEEE Trans. Pattern Anal. Mach. Intell. 2005;27(11):1832–1837. https://doi.org/10.1109/TPAMI.2005.220.
13. Zhu C, Yang J. Vision Based Hand Gesture Recognition Using 3D Shape Context. IEEE/CAA J. Autom. Sin. 2021;8(9):1600–1613. https://doi.org/10.1109/JAS.2019.1911534.
14. Xu C, Liu J, Tang X. 2D Shape Matching by Contour Flexibility. *IEEE Trans Pattern Anal Mach Intell*. 2009;31(1). https://doi.org/10.1109/TPAMI.2008.199.
15. Xu G, Li C. Plant leaf classification and retrieval using multiscale shape descriptor. J Eng. 2021;467–475. https://doi.org/10.1049/tje2.12050.
16. Yang C, Fang L, Fei B, Yu Q, Wei H. Multi-level contour combination features for shape recognition. Comput Vis Image Underst. 2023;229. https://doi.org/10.1016/j.cviu.2023.103650.
17. Yan T, Liu Y, Wei D, Sun X, Liu Q. Shape analysis of sand particles based on Fourier descriptors. Environ Sci Pollut Res. 2023;30:62803–62814. https://doi.org/10.1007/s11356-023-26388-5.
18. Yen Wu W. Shape Recognition Using Segmenting and String Matching. Asian J Appl Sci. 2022;10:528–536. https://doi.org/10.24203/ajas.v10i6.7138.
19. Li Z, Guo B, Meng. Fast Shape Recognition Method Using Feature Richness Based on the Walking Minimum Bounding Rectangle over an Occluded Remote Sensing Target. MDPI. *Remote Sens.* 2022;14(22):1–21. https://doi.org/10.3390/rs14225845.
20. González J, Navarro M, Hernández H. Shape Descriptor Based on Curvature. OALib J. 2022;9(3). https://doi.org/10.4236/oalib.1108422.
21. Mirehi N, Tahmasbi M, Targhi A. New graph-based features for shape recognition. Soft Computing: Methodologies and Applications. 2021;25:7577–7592. https://doi.org/10.1007/s00500-021-05716-2.
22. Paramarthalingam A, Thankanadar M. Extraction of compact boundary normalization based geometric descriptors for affine invariant shape retrieval. WILEY. IET Image Process. 2020;15:1093–1104. https://doi.org/10.1049/ipr2.12088.
23. Aswathi AS, India K. Curvature Bag of Words Model for Shape Based Recognition and Image Retrieval. *Int J Innov Res Sci Eng Technol*. 2020;9:5469–5505.
24. Zheng Y, Meng F, Liu J, Guo B, Song Y, Zhang X, *et al.* Fourier Transform to Group Feature on Generated Coarser Contours for Fast 2D Shape Matching. IEEE Access. 2020;8:90141–90152. https://doi.org/10.1109/ACCESS.2020.2994234.
25. Ahmed M, Aradhya M. 2D Shape Recognition and Retrieval Using Shape Contour Based on the 8-Neighborhood Patterns Matching Technique. *Int J. Softw Eng.* 2020;10:49–60. https://doi.org/10.4018/IJSE.2019070104.
26. Zheng Y, Guo B, Li C, Yan Y. A Weighted Fourier and Wavelet-Like Shape Descriptor Based on IDSC for Object Recognition. MDPI. Symmetry. 2019;11:693. https://doi.org/10.3390/sym11050693.
27. Zheng Y, Guo B, Chen Z, Li C. A Fourier Descriptor of 2D Shapes Based on Multiscale Centroid Contour Distances Used in Object Recognition in Remote Sensing Images. MDPI, Sensors. 2019;19:3–19. https://doi.org/10.3390/s19030486.
28. Rababaah A, Rabaa'I A. Geometric 2D Shapes Recognition with Polar Signature Characterization and Template Matching. Review of Business and Technology Research. 2017;14(2):7–12.
29. Shen Y, Ai T, He Y. A New Approach to Line Simplification Based on Image Processing: A Case Study of Water Area Boundaries. ISPRS Int J Geo-Inf. 2018;7(2):1–25. https://doi.org/10.3390/ijgi7020041.
30. Visvalingam M, Whyatt'J D. The Douglas-Peucker Algorithm for Line Simplification: Re-evaluation through Visualization. *Comput Graph Forum*. 2007;9(3):213–225. https://doi.org/10.1111/j.1467-8659.1990.tb00398.x.
31. Kolanowsk B, Augustyniak J, Latos D. Cartographic Line Generalization Based on Radius of Curvature Analysis. Geo-Inf. 2018;7(12):1–21. https://doi.org/10.3390/ijgi7120477.
32. Ekdemir S. Efficient Implementation of Polyline Simplification for Large Datasets and Usability Evaluation. Master. Sweden: Uppsala University, Disciplinary Domain of Science and Technology, Mathematics and Computer Science, Department of Information Technology. 2011:1–53.
33. Wang W, Yang W, Liu Y, Sun R, Hu J, Yang L. Segmented Douglas-Peucker Algorithm Based on the Node Importance. KSII Trans. Internet Inf. Syst. 2020;1562:1577. http://doi.org/10.3837/tiis.2020.04.009.
34. Li Z, Xin Q, Sun Y, Cao M. A Deep Learning-Based Framework for Automated Extraction of Building Footprint Polygons from Very High-Resolution Aerial Imagery. MDPI, Remote Sens. 2021;13(18):1–25. https://doi.org/10.3390/rs13183630.
35. Jung LIM S. A Chain-Based Wireless Sensor Network Model Using the Douglas-Peucker Algorithm in the IoT Environment. Tehnički Vjesnik. 2021;28(6):1825–1832. https://doi.org/10.17559/TV-20200916075229.
36. LIU j, Li H, Yang Z, WU K, Liu Y, Liu A. Adaptive Douglas-Peucker Algorithm With Automatic Thresholding for AIS-Based Vessel Trajectory Compression. IEEE Access. 2019;7:150677–150692. https://doi.org/10.1109/ACCESS.2019.2947111.
37. Zambrano G, Sentí V. Comparison analysis on noise reduction in GPS trajectories simplification. 19th LACCEI. 2021:1–6. https://doi.org/10.18687/LACCEI2021.1.1.96.
38. Tienaah T, Stefanakis E, Coleman D. Contextual Douglas-Peucker Simplification. *Geomatica*. 2015;69:327–338. https://doi.org/10.5623/cig2015-306.
39. Mirvahabi S, Abbaspour R, Claramunt C. A Flexible Trajectory Compression Algorithm for Multi-Modal Transportation. ISPRS. Annals of the Photogrammetry, Remote Sensing and

Spatial Information Sciences. 2023;X-4/W1-2022:501–508. https://doi.org/10.5194/isprs-annals-X-4-W1-2022-501-2023.

40. Liu B, Liu X, Li D, Shi Y, Fernandez G, Wang Y. A Vector Line Simplification Algorithm Based on the Douglas–Peucker Algorithm, Monotonic Chains and Dichotomy. MDPI. Geo-Inf. 2020;9(4):1–14. https://doi.org/10.3390/ijgi9040251.
41. Devikanniga1 D, Ramu1 A, Haldorai A. Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crows Search Algorithm. EAI Endorsed Trans Energy Web. 2020;7(29):1–10. http://dx.doi.org/10.4108/eai.13-7-2018.164177.
42. Deiss L, Margenotb A, Culman S, Demyan M. Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. Geoderma. 2020;365:1–12. http://dx.doi.org/10.1016/j.geoderma.2020.114227.
43. Yassin B, Mohamed C, Yassine A. A Nonlinear Support Vector Machine Analysis Using Kernel Functions for Nature and Medicine. EDP Sciences. https://doi.org/10.1051/e3sconf/202131901103.
44. Mahdi G. A Modified Support Vector Machine Classifiers Using Stochastic Gradient Descent with Application to Leukemia Cancer Type Dataset. Baghdad Sci J. 2020;17(4):1255–1266. https://orcid.org/0000-0003-4870-4034.
45. Jun Z. The Development and Application of Support Vector Machine. IOP. J Phys. 2021;1748:1742–1748. https://doi.org/10.1088/1742-6596/1748/5/052006.
46. Khanday A, Khan Q, Rabani S. Detecting Textual Propaganda Using Machine Learning Techniques. Baghdad Sci J. 2021;18(1):199–209. https://doi.org/10.21123/bsj.2021.18.1.0199.
47. Kumar S, Samriya JK, Yadav AS, Kumar M. To improve scalability with Boolean matrix using efficient gossip failure detection and consensus algorithm for PeerSim simulator in IoT environmentInt. J Inf. Technol. 2022;14(5):2297–2307. https://doi.org/10.1007/s41870-022-00989-8.

اسلوب جديد لتميز انماط الاشكال اعتمادا على صفات محيط الشكل المستخرجة بواسطة خوارزمية تبسيط الخط

علي عادل سعيد، رحيم عكلة، شيماء حميد شاكر

قسم علوم الحاسوب، الجامعة التكنولوجية ،بغداد،العراق

الخلاصة

تعتبر عملية تمييز الانماط الاشكال احدى المهام الرئيسة في التطبيقات المتعلقة بروئية الماكنة وذلك اعتمادا على التقنيات المعتمدة في استخلاص صفات الشكل والتي تكون مصنفة على اساس المحيط او المنطقة الداخلية للشكل.في هذا البحث تم التطرق الى اقتراح خوارزمية لاستخلاص متجة صفات الشكل اعتمادا على مجموعة النقاط التقديرية لمحيط الشكل اعتماداى على خوارزمية (دوكلاس بوكر) عن طريق تكرار استدعاء الخوارزمية لعدد غير محدد من المرات بعد الابتداء بحد سماحية التقارب وزيادة قيمته بصور تصاعدية مع كل تكرار لغاية الحصول على العدد المطلوب. ومقابل كل عدد لناقط التقارب يتم توليد متجة الصفات اعتمدا على مصفوفة المسافة واجراء فحص تميز نمط الشكل مع انماط الاشكال المخزونة في قاعدة الاشكال .لاجل بيان الافتراض القائم على ان هل هنالك عدد امثل لنقاط التقارب المستخرجة من خوارزية التقارب تكون مثلى لتحقيق اعلى نسبة للتمييز.تم اجراء الاختبار على مجموعة الاشكال الاختبارية (MPEG-7) باستخدام المصنف متجة الماكنة الداعم (SVM) وتبين صحة الفرضية بوجود عدد محدد لنقاط القارب يحقق اعلى نسبة تمييز بنسبة 0.961

الكلمات المفتاحية: دوكلاس بوكر، تبسيط الخط، تمييز الانماط، تمييز الاشكال، SVM.