Postprint: A Fast Shape Matching Algorithm Based on L1 Norm

Authors: Wang Jianghui, Wu Xiaojun

Date: 2018-05-20T00:00:00+00:00

Abstract

To address the issues of time-consuming histogram matching and poor engineering applicability in shape similarity measurement for inner-distance shape context (IDSC) and contours points distribution histogram (CPDH), a method using EMD-L1 to measure the distance of contour feature histograms is proposed. EMD-L1 integrates the L1 norm based on the original Earth Mover' s Distance (EMD). By replacing the ground distance calculation method, it reduces the variables in the objective function, accelerates histogram matching speed, and enables rapid shape matching while maintaining favorable retrieval performance. Simulation experimental results on shape datasets demonstrate that the proposed method can effectively perform shape recognition and retrieval on datasets, and its matching speed on the MNIST dataset outperforms other algorithms.

Full Text

Fast Shape Matching Algorithm Based on L1 Norm

Wang Jianghui, Wu Xiaojun
School of IoT Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract

To address the problems of long histogram matching time and poor engineering applicability associated with Inner-Distance Shape Context (IDSC) and Contour Points Distribution Histogram (CPDH) in shape similarity measurement, this paper proposes using EMD-L1 to measure the distance between contour feature histograms. EMD-L1 integrates the L1 norm into the original Earth Mover' s Distance (EMD) by replacing the ground distance calculation method, which reduces the variables in the objective function and accelerates histogram matching speed, enabling rapid shape matching while maintaining good retrieval performance. Simulation experiments on shape datasets demonstrate that the proposed method can effectively perform shape recognition and retrieval, and its matching speed on the MNIST dataset outperforms other algorithms.

Keywords: inner-distance shape context; contour points distribution histogram; earth mover’ s distance (EMD); L1 norm; shape retrieval

---

0 Introduction

Shape is a crucial feature of objects that conveys deeper visual information compared to appearance characteristics such as texture and grayscale. Shape matching finds applications in numerous computer vision domains, including object recognition and image retrieval, where shape serves as a basis for classification and identification. Shape retrieval can be categorized into 2D and 3D shape retrieval. While deep learning-based 3D shape retrieval algorithms have gained significant attention from researchers in recent years, 2D shape retrieval remains a representative problem in computer vision.

The 2D shape matching process primarily consists of two components: first, designing a shape descriptor to represent the shape through some method; then calculating the distance between descriptors of different shapes to measure their differences and complete matching. This approach can be divided into three categories based on shape representation: shape point sets, shape boundary contours, and shape internal skeletons. Contour-based methods incorporate ordering information of points, while skeletons contain geometric features and topological structures of shapes.

Among contour-based methods are multi-scale approaches such as Triangle-Area Representation (TAR), Multiscale Convexity (MCC), and Multiscale Integral Invariants. These methods construct shape features at multiple scales to enrich feature information. Shape Context (SC) is a classic method based on the relative positional relationships of contour points. This algorithm focuses on the spatial distribution relationship between a particular contour point and all other points, but suffers from poor noise resistance and time-consuming histogram distance calculation using Thin Plate Spline (TPS) functions.

Ling proposed the Inner-Distance Shape Context (IDSC) method, which replaces Euclidean distance with inner distance between contour points to measure differences. This approach provides good representation for non-rigid object contours and uses dynamic programming to measure distances between target shapes, though its high computational complexity limits practical applicability. Shu and Wu proposed a 2D shape descriptor called Contour Points Distribution Histogram (CPDH), which uses the minimum enclosing circle of a target shape as the feature extraction region and constructs a contour point distribution histogram in polar coordinates centered at the circle’ s origin. CPDH employs Earth Mover’ s Distance (EMD) to measure histogram distances, but its focus on local contour features leads to poor performance on datasets with large deformations, unstable retrieval, and high EMD algorithm time complexity that makes shape matching too slow for practical use.

This paper proposes using EMD-L1 to measure the similarity between two shape contour feature histograms generated by contour descriptors. EMD-L1 leverages the L1 norm (Manhattan distance) as a new ground distance on top of the original EMD. The improved EMD-L1 reduces computational complexity and accelerates shape matching speed by decreasing the number of variables and constraints in the objective function, without weakening the discriminative power of shape descriptors. Additionally, the L1 norm offers advantages in feature extraction such as denoising and reducing outliers, making shape descriptors more robust. Experimental analysis across various datasets demonstrates that the improved algorithm effectively maintains the good retrieval performance of IDSC and CPDH while improving matching speed and enhancing engineering applicability.

1 Fundamentals of CPDH and IDSC Algorithms

Contour features are the most essential characteristics of images. CPDH is a target feature descriptor proposed based on contour point distribution, while IDSC is a shape descriptor with good representation capability based on spatial distribution relationships between sample points and inner distances.

1.1 Constructing the CPDH Descriptor

First, the standard Canny operator is used to extract the target contour from the image. However, since target shape contours contain too many points, an equal-interval sampling method is applied to reduce computational load and meet real-time requirements, obtaining partial contour points. The number of points is denoted by n, which describes the object' s shape.

As shown in [Figure 1: see original paper], (a) represents the original image, (c) shows the sampled contour image, and (d) displays the minimum enclosing circle of the target shape. Using the contour point set P as the target region, we construct its minimum enclosing circle C and establish a polar coordinate system with the circle center (Cx, Cy) as the origin. Within this minimum enclosing circle, a sector grid model is built using concentric circles to 统计轮廓点的分布特征, as illustrated in [Figure 2: see original paper].

For each sector region Ri, we obtain a triple (i, i, ni), where i is the radius of the i-th concentric circle, i is the angle between a sector edge of region Ri and the horizontal direction (rotating clockwise), and ni is the number of points falling within region Ri. Thus, an image I can be described as a histogram H consisting of R triples. Specific details of CPDH description can be found in reference [7].

1.2 IDSC Shape Descriptor

First, the second-order gradient Canny operator is used to extract partial contour points of the target shape. For any sampling point, it can be represented by vectors constructed from the remaining n-1 points and the point itself. Therefore, the entire target shape can be represented by n×(n-1) vectors, which contain both distance and orientation information of contour points within the whole shape. The orientation parameter represents the number of equal divisions of a circle’ s circumference, while the distance parameter represents the number of equal divisions of the radius. The resulting vector set is then represented using a shape statistical histogram.

Inner distance is defined as the length of the shortest path connecting two points m and n inside shape A, denoted as dA(m,n). The shortest path generally has two cases: if the line segment between points m and n lies completely within the target shape, the inner distance is equivalent to Euclidean distance; otherwise, the Floyd-Warshall shortest path algorithm is used to calculate the distance between each pair of sampling points. Combining the above components yields the IDSC of the target shape.

As shown in [Figure 3: see original paper], Figure 3(a) displays the extracted bat shape contour with four contour points marked, while Figure 3(b) shows the corresponding inner-distance shape context features for these marked points.

---

2 EMD Algorithm

Similarity measurement is a crucial 环节 that directly affects retrieval performance, and researchers have long 致力于研究如何用数值来有效地表示图像在特征上的相似程度. Rubner et al. [13] proposed the EMD algorithm, which is widely applied to transportation problems in linear programming. Intuitively, this paper views two compared images as two different distributions of image features in space –one as a pile of earth and the other as a pit. The EMD algorithm finds the optimal path to complete the earth transportation task, where the work done under this optimal path can be considered as the distance between the two distributions.

The algorithm can be formally described as the following linear programming problem: Let space S distribute m earth piles P = {(p1, wp1), ···, (pm, wpm)}, while simultaneously distributing n pits Q = {(q1, wq1), ···, (qn, wqn)}, where the j-th pit qj can accommodate earth mass wqj. Let D = [dij] represent the ground distance matrix. We need to find the optimal transportation scheme F = [fij], where fij represents the mass of earth transported from pile pi to pit qj. The minimum work for earth transportation is:

$$\text{Work}(P, Q, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}$$

This must satisfy the following four constraints: 1. Condition 1 limits that earth can only be transported from pile set P to pit set Q, not vice versa; 2. Condition 2 limits that the maximum mass of earth that pile pi can transport does not exceed wpi; 3. Condition 3 limits that the maximum mass of earth that pit qj can accommodate does not exceed wqj; 4. Condition 4 indicates transporting as much earth as possible.

Once this transportation problem is solved, the optimal transportation scheme F is found, and the EMD distance can be calculated using the final normalized form:

$$\mathrm{EMD}(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}}$$

---

3 EMD-L1 Metric

CPDH uses the original EMD to measure similarity between contour feature histograms, but it suffers from high computational complexity and weak robustness to 微小形变. IDSC uses chi-square 检验 to calculate shape histogram similarity to obtain a cost matrix, then applies dynamic programming to find the optimal match. Although these two measurement algorithms achieve high matching accuracy, their excessive computation time is unfavorable for engineering applications.

To address this deficiency, this paper adopts EMD-L1 as the measurement criterion. The core of the EMD-L1 algorithm lies in replacing the ground distance of the original EMD with the L1 norm (Manhattan distance). The L1 norm (Manhattan distance) refers to the sum of absolute values of elements in a vector. Therefore, the newly defined ground distance is calculated as in equation (7). By replacing the original ground distance calculation method with the L1 norm, the number of variables in the objective function is reduced, accelerating histogram matching speed. Since the L1 norm demonstrates better robustness to noise and outliers than the L2 norm, it is selected for this application.

Let (i,j) and (k,l) represent blocks in the contour feature histograms of the target and query shapes, respectively. Given two m×n 2D histograms P and Q generated by CPDH, each containing N blocks, we define:

a) The index set of histogram blocks: = {(i,j) | 1 i m, 1 j n}, where (i,j) represents a block in the histogram.

b) The set of flows between blocks: = {(i,j,k,l) | (i,j) , (k,l) }.

c) From (a), we obtain the definitions of histograms P and Q: P = {pij | (i,j) }, Q = {qij | (i,j) }.

Based on these three definitions, we redefine the EMD between histograms P and Q:

$$\text{EMD}(P,Q) = \min_F \sum_{(i,j,k,l)\in\nu} d_{i,j;k,l} f_{i,j;k,l}$$

where F represents the flow from histogram P to Q; f_{i,j;k,l} represents the flow from block (i,j) to (k,l); and the ground distance d_{i,j;k,l} can be calculated using equation (7).

To further optimize the objective function, is divided into three subsets, each corresponding to a flow under different circumstances:

 = {(i,j,i,j) }, representing flows between blocks at the same histogram position, named s-flows.

 = {(i,j,k,l) | |i-k| + |j-l| = 1}, representing flows between adjacent blocks in the histogram, named n-flows.

 = {(i,j,k,l) | |i-k| + |j-l| > 1}, representing the remaining flows, named f-flows.

According to the above definitions, f-flows can be divided into several n-flows because the L1 distance on integer histogram blocks can form a shortest path system. For example, given an f-flow, the L1 ground distance can be decomposed as:

$$d_{i,j;k,l} = \sum_{x=i}^{k-1} d_{x,j;x+1,j} + \sum_{y=j}^{l-1} d_{k,y;k,y+1}$$

In summary, the EMD-L1 distance between contour histograms under CPDH for target shape P and query shape Q is:

$$\text{EMD-L1}(P,Q) = \min_G \sum_{(i,j,k,l)\in\nu_1} g_{i,j;k,l}$$

subject to:

$$\begin{cases} g_{i,j;k,l} \geq 0, & \forall(i,j,k,l)\in\nu_1 \\ \sum_{(k,l)\in\eta} g_{i,j;k,l} - \sum_{(k,l)\in\eta} g_{k,l;i,j} = b_{ij}, & \forall(i,j)\in\eta \end{cases}$$

where b_{ij} = p_{ij} - q_{ij} represents the difference in contour point counts between the two histograms at block (i,j).

Compared with the original EMD, the objective function (11) and constraint (12) reduce the number of variables and constraints. Additionally, based on the characteristics of L1 distance, there is no need to calculate ground distances between blocks, significantly reducing algorithmic computational complexity.

4 Algorithm Time Complexity Analysis

The optimization problem in the improved algorithm reduces the number of variables in the objective function from O(N^2) to O(N), where N represents the number of blocks divided in the enclosing circle. EMD-L1 utilizes the L1 norm to calculate ground distance, substantially improving histogram matching speed. The original EMD algorithm has a complexity of O(N^3 log N), while the optimized algorithm complexity is O(N). Experimental verification confirms that the improved algorithm effectively enhances shape matching speed.

---

5 Experimental Results and Analysis

To verify the effectiveness of the proposed method, experiments were conducted on different datasets. During testing, each sample in the dataset was treated as a query shape, with the remaining samples as target shapes. Shape distance results were sorted in ascending order. If the matching results showed that the query shape and target shape belonged to the same category, the match was considered correct.

5.1 Kimia-25 and Kimia-99 Shape Datasets

The Kimia-25 database [14] contains 25 samples across 6 categories, with examples shown in Figure 4: see original paper. The Kimia-99 database [15] contains 99 samples across 9 categories, with 11 samples per category, as shown in Figure 4: see original paper.

In all experiments, to ensure fair comparison with other methods, the number of sampling points for each shape contour was set to N = 100. During testing on Kimia-25 and Kimia-99 databases, the target shape' s minimum enclosing circle was divided into 12 equal angular segments and 5 equal radial segments, resulting in 5 × 12 = 60 region blocks.

The simulation results on the Kimia-25 database are presented in , which lists retrieval results under different methods on the Kimia-25 dataset, showing the number of correctly classified results among the 1st to 3rd most similar target shapes to the query shape. The results indicate that the proposed algorithm outperforms the classic SC method, primarily because SC' s local histogram construction suffers from poor discriminative ability when the radius is too short, and includes excessive noise when the radius is too long. However, the proposed algorithm' s retrieval performance is weaker than CPDH+EMD and IDSC+DP. Analysis reveals that replacing the original EMD ground distance with L1 distance ignores spatial information between contour points, leading to slightly decreased retrieval performance.

The simulation results on the Kimia-99 database are shown in , presenting the number of correctly classified results among the 1st to 10th most similar target shapes and the total count. The table demonstrates that the proposed algorithm' s retrieval performance is superior to SC but weaker than CPDH+EMD. However, when retrieving the 8th to 10th most similar classes, the proposed algorithm is more stable than the other two algorithms, returning more valid shapes. This is because the L1 distance provides better robustness to 微小形变. Additionally, the improved IDSC algorithm' s retrieval performance is significantly better than SC and CPDH+EMD, proving that EMD-L1 can well preserve IDSC' s shape description capability.

5.2 Articulated Shape Dataset

The Articulated dataset [10] contains 8 classes with a total of 40 images. Each class has 5 images with varying degrees of deformation, as shown in [Figure 6: see original paper]. This dataset is used to test shape articulation, which represents a special case of deformation, and can also 检验算法对形变的鲁棒性.

This experiment employs cross-validation, using chi-square 检验 and EMD-L1 to measure similarity between histograms generated by the same shape descriptor. The number of correctly retrieved target shapes among the 1st to 4th most similar shapes returned is used as the evaluation criterion. Analysis of experimental data shows that comparing the first and second groups of experiments, EMD-L1 performs better than chi-square 检验 on datasets with obvious deformation, demonstrating greater robustness. Comparing the second and third groups, IDSC' s shape discriminative ability is superior to SC. Comparing the first and fifth groups, CPDH' s ability to describe shape deformation is better than SC because IDSC' s shape description includes relative positional relationships between contour points, while SC has poor noise resistance.

5.3 MPEG-7 Dataset

Based on the proposed algorithm, experiments were conducted on the MPEG-7 dataset [16], which contains 1400 shape silhouettes evenly divided into 70 categories with 20 samples per category.

5.3.1 Comparison of CPDH+EMD and CPDH+EMD-L1 presents a comparison between CPDH+EMD and CPDH+EMD-L1 algorithms. This experiment was conducted on a subset of MPEG-7 Set B, which includes 216 shape images divided into 18 categories with 12 images per category. The results show that although CPDH+EMD returns more valid shapes in total than CPDH+EMD-L1, analyzing retrieval results from the 8th to 11th classes—especially the 10th to 11th classes—reveals that EMD' s returned shape count decreases rapidly while EMD-L1 decreases steadily, with far more shapes returned in the 11th class than the original algorithm. This indicates that CPDH based on EMD-L1 has superior robustness to deformation when the number of classes is large. However, due to the L1 norm' s characteristic of ignoring spatial information between contour points, the overall effect is slightly inferior to the original algorithm.

5.3.2 Comparison of IDSC+EMD-L1 with Other Methods This experiment uses the MPEG-7 dataset and employs the Bullseye [17] method to determine retrieval accuracy: each query shape is compared with all shapes in the dataset, and the 40 most similar shapes are retrieved. The number of shapes belonging to the same category as the query shape among these 40 is counted, and the corresponding recall rate is calculated. shows that EMD-L1 improves IDSC' s retrieval performance and surpasses most shape retrieval algorithms, proving the effectiveness of the proposed algorithm.

6 Rotation Invariance Test

To verify the rotation invariance of IDSC+EMD-L1, this experiment randomly selected 10 images from the MPEG-7 dataset as reference images. Each image was then rotated clockwise and counterclockwise by 5°, 15°, 30°, 45°, 60°, and 75° to generate 12 new images. Including the 10 reference images, the rotation dataset contains 130 images. lists the retrieval results using the 10 reference images as queries. With 118 images retrieved correctly and only 12 retrieval errors, the improved algorithm maintains good rotation invariance.

7 Algorithm Speed Test

This experiment compares two groups of algorithms. The first group: CPDH+EMD-L1, CPDH+EMD, and SC. The second group: IDSC+DP, IDSC+EMD-L1, and SC. The dataset used is the MNIST [22] handwritten digit image training set, which contains 10,000 images of digits 0-9. 1,000 images were randomly selected for the experiment (digit shapes range from low to high complexity, making the calculated recognition speed more general). The results are shown in [Figure 6: see original paper] and [Figure 7: see original paper]. Experiments prove that the improved algorithm' s running time is significantly better than the original algorithms and SC, making it more suitable for engineering applications.

8 Conclusion

This paper replaces EMD and DP algorithms in CPDH and IDSC with EMD-L1 as the measurement criterion for contour point histograms. The main advantage lies in the L1 norm (Manhattan distance) providing better robustness to deformation and reducing algorithmic computational complexity by decreasing the number of variables in the objective function. Experimental results demonstrate that the improved algorithm maintains good retrieval capability and accelerates shape matching speed. However, the L1 norm's characteristic of ignoring spatial information between contour points leads to slightly decreased retrieval performance. Given the recent surge in deep learning, 3D shape retrieval based on neural networks will be the author' s next research direction.

References

[1] Zhang D, Lu G. Review of shape representation and description techniques [J]. Pattern Recognition, 2004, 37(1): 1-19.
[2] Yang M, Kpalma K, Ronsin J. A survey of shape feature extraction techniques [J]. DOI: 10.5772//6237, 2008.
[3] Bai S, Bai X, Zhou Z, et al. Gift: A real-time and scalable 3d shape search engine [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 5023-5032.
[4] Zhu Z, Wang X, Bai S, et al. Deep learning representation using autoencoder for 3D shape retrieval [J]. Neurocomputing, 2016, 204: 41-50.
[5] 周瑜, 刘俊涛, 白翔. 形状匹配方法研究与展望 [J]. 自动化学报, 2012, 38(6): 889-910.
[6] Alajlan N, EI Rube, Kamelb M S, Freeman G H. Shape retrieval using triangle-area representation and dynamic space warping [J]. Pattern Recognition, 2007, 40(7): 1911-1920.
[7] Adamek T, O' Connor N E. A multiscale representation method for nonrigid shapes with a single closed contour [J]. IEEE Trans on Circuits and Systems for Video Technology, 2004, 14(5): 742-753.
[8] Hong B W, Soatto S. Shape Matching using Multiscale Integral Invariants [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2015, 37(1): [9] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2002, 24(4): 509-522.
[10] Ling H, Jacobs D W. Shape classification using the inner-distance [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(2): 286-99.
[11] Shu Xin, Wu Xiaojun. A novel contour descriptor for 2D shape matching and its application to image retrieval [J]. Image and Vision Computing, 2011, 29(4): 286-294.
[12] Ling, Haibin, and Kazunori Okada. An efficient earth mover' s distance algorithm for robust histogram comparison. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(5): 840-853.
[13] Yossi R, Guibas L J, Tomasi C. The earth mover' s distance multidimensional scaling and color-based image retrieval [C]// Proc. of ARPA Image Understanding Workshop. 1997: 661-668.
[14] Sharvit D, Chan J, Tek H, et al. Symmetry-based indexing of image databases [C]// Proc of IEEE Workshop on Content-Based Access of Image and Video Libraries. 1998: 56-62.
[15] Sebastian T B, Klein P N, Kimia B B. Recognition of shapes by editing their shock graphs [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2004, 26(5): 550-571.
[16] Attalla E, Siy P. Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching [J]. Pattern

Recognition, 2005, 38(12): 2229-2241.
[17] Latecki L J, Lakamper R, Eckhardt T. Shape descriptors for non-rigid shapes with a single closed contour [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2000: 424-429.
[18] Mokhtarian F, Abbasi S, Kittler J. Robust and efficient shape indexing through curvature scale space [C]// Proc of British Machine and Vision Conference. 1996, 96: 53-62.
[19] Sebastian T B, Klein P N, Kimia B B. On aligning curves [J]. IEEE Trans on pattern analysis and machine intelligence, 2003, 25(1): 116-125.
[20] Grigorescu C, Petkov N. Distance sets for shape filters and shape recognition [J]. IEEE Trans on Image Processing, 2003, 12(10): 1274-1286.
[21] Tu Zhuowen, Yuille A L. Shape matching and recognition–using generative models and informative features [C]// Proc of European Conference on Computer Vision. Berlin: Springer, 2004: 195-209.
[22] Deng Li. The MNIST database of handwritten digit images for machine learning research [best of the web] [J]. IEEE Signal Processing Magazine, 2012, 29(6): 141-142.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.