Image and Vision Computing 27 (2009) 1097–1107

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Decomposition of two-dimensional shapes for efficient retrieval

Cecilia Di Ruberto [a,*], Luigi Cinque [b]

[a] Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124 Cagliari, Italy
[b] Department of Computer Science, University of Rome "La Sapienza", Rome, Italy

ARTICLE INFO

Article history:
Received 5 October 2006
Received in revised form 11 July 2008
Accepted 9 October 2008

Keywords:
Shape decomposition
Morphological skeleton
Shape representation
Shape retrieval
Shape matching
Object recognition

ABSTRACT

This paper presents a novel approach to address the problem of generic 2D shape recognition. We propose a morphological method to decompose a binary shape into entities in correspondence with their protrusions. Each entity is associated with a set of perceptual features that can be used in indexing into image databases. The matching process, based on the softassign algorithm, has produced encouraging results, showing the potential of the developed method in a variety of computer vision and pattern recognition domains. The results demonstrate its robustness in the presence of scale, reflection and rotation transformations and prove the ability to handle noise and articulated structures. In order to increase efficiency, the retrieval process is applied after a coarse scale grouping of objects, without sacrificing effectiveness and allowing indexing into large shape databases.

1. Introduction

There has been large work in the area of shape comparison for object recognition and indexing. A number of approaches consider an object as a single feature vector in a high-dimensional space [47]. Unfortunately they are not able to predict entry-level categories because there is no abstraction from image data to a model. In [32] shapes are represented as modal deformations of a prototype object. Shape similarity is evaluated as distance between mode vectors. Del Bimbo and Pala [7] have computed shape similarity by minimizing a cost function which considers the amount of deformation of the sketch and the degree of matching between the deformed sketch and the query shape. There are also systems which define global attributes and use vector model for indexing in image databases. In [12] a set of 22 global shape features including circularity, area, major axis orientation and a set of algebraic moments are defined and indexing is performed on a lower dimensional space by using R^*-tree. In [20] shapes are represented by two sets of global features: a 72 bins histogram of the shape edge direction and 7 invariant moments. Shape similarity is evaluated as a weighted sum of the Euclidean distance between the histograms. The authors do not propose indexing. Of course the use of global features shows limitations in modelling perceptual aspects of shapes. On the contrary, local feature-based methods present a more effective performance. There are methods which represent each object as a collection of line segments, curves, corners, regions [11,13,35]. In [40] the author proposes an adaptive rectangular decomposition, where a binary object is decomposed into a union of rectangles. A dynamic programming is used to select the dimensions and locations of the rectangles so that the total number of rectangles required to represent the object is minimized. The author shows that this decomposition can represent objects with fewer points than the morphological skeleton. Another important result is that by incorporating a more sophisticated search technique, the author shows that a more efficient representation may be found. As a drawback the number of points required to represent the object can influence the computational cost. In [33] shapes are represented as a set of segments between two consecutive inflection points. Segments are considered at different levels of shape resolution and matching is achieved by minimizing a cost function. The success of these methods depends strongly on the choice of the features and their extraction. Moreover, such methods show that geometric variations can be too selective since global information could be lost in local details. The main task in all pattern recognition problems is to define a suitable perceptual model and automatically extract effective visual information to model as features vectors. In this way the shape matching process is reduced to evaluate distances between points in a multidimensional feature space. Since the use of global features presents limitations in modelling perceptual aspects of shape and poor performance in the computation of similarity with partially occluded shapes, we propose to decompose the shape boundary into perceptually significant entities with associated local boundary features. A shape is then represented by a collection of boundary feature vectors. In partitioning a shape boundary we have combined the use of the thinned representation of a shape with the distance function in order to first decompose the shape into different

* Corresponding author. Tel.: +39 070 6758507; fax: +39 070 6758511.
E-mail address: dirubert@unica.it (C. Di Ruberto).

0262-8856/$ - see front matter
doi:10.1016/j.imavis.2008.10.009

regions and then to achieve the boundary segmentation. After decomposing the shape boundary into entities, with each entity, we associate geometrical features used to index the original object in a database of shapes by minimizing a distance function. In [3] Berretti et al. propose a similar approach by using the curvature zero-crossing points from a Gaussian smoothed boundary to obtain the primitives, called tokens. With each token they associate its maximun curvature and its orientation as features. The token similarity is a weighted Euclidean distance. The method is not rotation invariant because of the curve orientation. For indexing the tokens into a feature database they use an M-tree structure. For the N tokens of the query shape they first search the similar tokens by visiting the index tree N times. The set of retrieved tokens with the same shape identifier represents a potential similar shape. Then they match the query shape and the potential similar shape using a model-by-model matching algorithm which is the best match between the tokens of the two shapes. The tokens matching involves thresholding which is ad hoc or empirical. Quantitative retrieval performance and retrieval efficiency are reported only for classical painted images. Also, the tree is visited a number of times in the shape matching so it is not clear whether the indexing is better than model-by-model indexing or not. Another comparable approach has been proposed by Latecki et al. in [26]. They present a shape similarity measure based on correspondence of boundary parts that are visually significant. They use a discrete evolution method as a prefilter for shape comparison in order to decompose a shape into visual parts. In [29] Mokhtarian et al. propose to use the maxima of curvature zero-crossing contours of curvature scale space image to represent the shapes of objects boundary contours [30]. The matching algorithm is based on global boundary information. The method is robust with respect to noise, scale and orientation changes of the object but it is sensitive to occlusion.

The remainder of this paper is organized as follows. Section 2 contains an overview of the proposed approach. In Section 3 the idea to use the morphological skeleton for shape partitioning is presented. Section 4 describes how to use local boundary features for an effective shape indexing. In Section 5 the shape matching algorithm is discussed. The evaluation of the method, indexing and a comparative analysis to similar approaches are presented in Section 6.

2. Overview of the method

Our approach for shape matching consists in decomposing a shape boundary into entities in correspondence with protrusions and modelling each entity according to a set of perceptually salient attributes. In this work we propose to use the skeleton of a shape in order to segment its boundary into parts by means of morphological operators as introduced in [8]. A skeleton can be partitioned into parts, each of them being a skeleton branch. From the skeleton parts we can also infer a meaningful decomposition of the object into regions, each of them being associated with one of the skeleton branches. We assign to each skeleton branch a region of the object by decomposing the object through the boundaries of the influence zones of the partitioned skeleton. Consequently from the partition of the object we can infer a segmentation of its boundary into entities, i.e. by intersecting the boundaries of the object regions with the shape boundary we obtain a shape boundary segmentation into entities. The partition occurs where the shape curvature presents sudden variations or, in other words, in correspondence to protrusions of the boundary curve. An entity can be modelled according to a set of perceptual significant attributes allowing an effective shape indexing. Two shapes are considered similar if they share entities with similar attributes, according to an appropriate distance measure. The entity distance is used to provide a measure of similarity between two entities. The shape distance is defined as a combination of entity distances and it is used to derive a global measure of shape similarity which fits human perception. Given two generic shapes, respectively, their distance is computed by combining the similarity of their entities according to a minimization procedure. In order to be useful, a shape retrieval system must have sufficiently high accuracy and speed to meet the needs of the user or application. For this reason we compare two shapes after a coarse scale grouping of the database images into families. A query shape is compared with all the images of a particular family and not with all the images in the database. Some recently published systems have employed complex shape matchers to achieve high accuracy, at the cost of relatively high pairwise matching times [37,44,46]. Our aim in this work was also to increase the speed of shape retrieval, without penalizing accuracy.

3. From morphological skeleton to shape decomposition

The skeleton represents a powerful tool for qualitative shape matching because it abbreviates information about the shape of the object and its topology, aiding synthesis and understanding [5,21]. In [9,10] we have introduced algorithms for detecting the main skeletons characteristic points (end points, junction points and curve points) based on a morphological approach. The detection of end points, junction points and curve points of medial axis is important for a structural description that captures the topological information embedded in the skeleton. Our approach for shape matching consists in decomposing a shape boundary into entities in correspondence with protrusions and in modelling each entity according to a set of perceptually salient attributes. In this work we propose to use the skeleton of a shape in order to segment its boundary into parts by means of morphological operators. In [9] we have proposed two different methods to identify end points and to detect junctions points from a skeleton using mathematical morphology. In this way a skeleton X of an object O can be considered as the union of the end points, the junctions and the curve points of X, i.e.

$$X = \mathrm{EndPoints}(X) \cup \mathrm{JunctionPoints}(X) \cup \mathrm{CurvePoints}(X).$$

As a consequence, a skeleton X can be partitioned into N parts $S = S_i(X), i = 1, \ldots, N$, each of them being a skeleton branch. From the skeleton parts we can also infer a meaningful decomposition of the object into regions, each of them being associated with one of the skeleton branches. With each skeleton branch $S_i(X)$ we associate a region $R_i(X)$ in the following way. Let S be the set of the skeleton branches computed by subtracting the junctions points from the skeleton X, i.e.

$$S = \bigcup_i S_i(X) = X \setminus \mathrm{JunctionPoints}(X).$$

We assign a region of the object O to each skeleton branch $S_i(X)$, by partitioning the object through the boundaries of the influence zones of S. We first detect the skeleton by influence zones of skeleton parts, $IZ(S)$ [38,43]. Indeed, the set of pixels of a binary image that are closer to a given connected component than any other connected component defines a partition of the image known as Voronöi diagram, each of them representing the influence zone of the considered connected component. There is a one-to-one correspondence between the set of connected components of a binary image and the set of its influence zones. The connected skeleton branches $S_i(X)$ are used as seeds (i.e. markers) for the splitting procedure. An Euclidean distance is used in the image decomposition. As in image processing field, the use of distance and markers permits to define influence zones and the boundary between influence zones is the skeleton by influence zone or *SKIZ*. The *SKIZ* principle is easily