CLASSIFICATION OF DIGITIZED CONTOURS REPRESENTED BY SIGNATURE

Krzyzak, A. and El buaeshi, H.
Department of Computer Science
Concordia University
Montreal, Quebec H3G 1M8

ABSTRACT

A classification of discrete contours via signatures is studied. Two algorithms to compute the signature are developed. In the first algorithm a multidimensional sorting is used. The second algorithm is based on a simple geometrical considerations. Two types of signature are considered –the length signature and the area signature. Statistical features based on Fourier descriptors are derived from the signatures. In classification the k-NN algorithm is used with k and the size of the feature vector chosen experimentally. The algorithms are tested on the handwritten, totally unconstrained characters from Suen's data base and recognition success rates of 91% and 93% are achieved for the length and area signature respectively.

KEYWORDS: curve signature, Fourier descriptors, contour classification.

INTRODUCTION

In many applications of pattern recognition and digital image processing the shape of a simply connected object is represented by its outer contour. Many shape recognition techniques deal with the entire object boundary, silhouette, intensity profile or range map. They include such methods as Fourier descriptors of the object boundary [2, 6, 7, 9, 10, 11, 15, 16, 18], moments of the silhouette [3, 14], and circular autoregressive models [8]. Among different techniques Fourier descriptors and curve signatures are distinguished by the invariance to the standard shape transformations such as scaling, rotation and translation. Some functions of Fourier descriptors are also invariant to mirror reflections and changes in the starting point [9, 15]. In this paper we studv the shape recognition problem using Fourier descriptors (FD's) derived from the curve signatures. This approach combines the simplicity of curve signatures with invariance of Fourier descriptors. We use the length signature proposed by O'Rourke [13] as well introducing an area signature. The latter is shown to be more robust with respect to the shape distortion. Efficient algorithms for computation of curve signatures are proposed and implemented. In shape classification features based on Fourier descriptors are used. Fourier descriptors are derived from the signature and are characterized by invariance to affine transformations, mirror reflections and changes in the starting point. We study several versions of Fourier descriptors and compare them. The shape representation methods proposed in this paper are tested in unconstrained handwritten numeral recognition. Basic components of the implemented recognition system are shown in Figure 1.

The feature extraction module consists of two sub-modules: signature module and Fourier descriptors module. The algorithm based on the geometrical

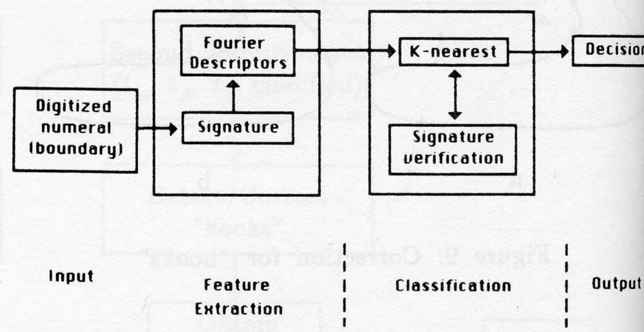

Figure 1. The basic components of the implemented character recognition system.

considerations is used to compute the length/area signature of the numeral. The Fourier descriptors module computes the FD's of the numerals from their signatures. The classification module is divided into two sub-modules as follows:

1) k-NN sub-module to find the k-nearest neighbor using branch and bound algorithm [4].
2) signature verification sub-module to separate distinct numerals with similar FD's.

In sub-module 1 FD's are used alone while in sub-module 2 classification is based on the signature.

CURVE SIGNATURE

Among feature extraction techniques used in shape recognition the most popular ones are: Fourier descriptors [7, 15, 18], boundary line encoding [10], polygonal approximation and directional and curvature feature extraction (see [17] for review). In this and the following sections feature extraction methods based on the signature and Fourier descriptors will be presented.

A signature of a plane curve is defined by O'Rourke [13]. Let Γ be a continuous, directed curve in the Euclidean plane and parametrized by its arc length t. The signature S(t) of Γ is the function which associates each point t of Γ with the length of Γ which is on or to the left of a tangent line at point t –Figure 2. We also consider the alternative version of the signature defined by taking the area to the left of a tangent line instead of the length to the left –Figure 3.

Among advantages of the curve signature we list its simplicity (the signature of a polygon is represented by the histogram) and invariance to shifts and rotations (as long as the starting point is maintained.) The main demerit of the signature is that it does not uniquely identify the curve it is derived from, except for rectilinear curves [13]. For instance all convex curves of unit length are mapped on the constant signature at one or zero