PROGRESSIVE IMAGE CODING FOR VISUAL SURVEILLANCE APPLICATIONS BASED ON STATISTICAL MORPHOLOGICAL SKELETON

G.L. Foresti, C.S. Regazzoni and A. Teschioni

Department of Biophysical and Electronic Engineering (DIBE), University of Genoa
Via all'Opera Pia 11A, 16145 Genova, Italy.
Phone +39-10-3532-792, Fax +39-10-3532-134
e-mail: forte@dibe.unige.it

ABSTRACT
This paper presents a new shape representation method for progressive image coding at very-low bit rate. A real application in a railway surveillance system for unattended level-crossings is considered. First, semantic information, e.g., classification of possible obstacles provided by a recognition subsystem, is sent to a remote control center; then, binary shape information is transmitted, in order to allow the remote operator to validate the alarm situation. Pictorial information can be required as a further step by the operator of the control center.

1 INTRODUCTION
Representation of shape information for coding and recognition purposes have been often considered as two separate problems. Image representation used within coding systems is evaluated in terms of compression rates and quality of reconstructed images. On the other hand, image representation employed in recognition systems aims at extracting features to be matched with available object models.

In recent years, the demand for services is increasing where both automatic recognition and coding are performed, so providing an impulse to research on new image representation methods providing a common basis to both problems. Remote evaluation of the presence of specific objects within a large database of images is a prototypical problem implying recognition and coding aspects. Retrieval of visual information in multimedia systems, as well as detection of alarm situations within surveillance systems (e.g., for public transports) are applicative examples of such a problem [1,2].

In [3], a image representation method for lossy binary image coding of noisy binary images in surveillance applications has been presented which is based on an approximation of the Statistical Morphological Skeleton (SMS) [4]. The approach proposed in [3] provides a noise-robust shape representation method which can be used for two purposes: (a) very low bit rate shape transmission; (b) automatic object identification by matching with a pre-defined set of models [5]. Noise-robustness concerns with both the representation itself and the shape reconstructed after transmission. In fact, obtained representations are shown to vary continuously with the slight changes of noise level on data. The representation introduced in [3] dealt only with shape characteristics of the object to be recognized and transmitted.

In this paper, we extend that representation in order to take into account also pictorial (i.e., luminosity) information. The extension is performed in such a way that luminosity information is separated from other information to be transmitted, i.e., semantic and shape information. First, semantic information can be sent, e.g., classification of the object provided by a decentralized automatic recognition subsystem; then, binary shape information can be transmitted, in order to allow the remote user to validate the alarm signal. Pictorial information can be required as a further step by the operator of the control center. This choice allows the system the capability of transmitting information in a progressive fashion, depending on the user requirements.

2 SYSTEM DESCRIPTION
The method presented in this paper consisted of four main steps: (a) extraction of the Statistical Morphological Skeleton (SMS); (b) geometrical approximation of the skeleton points by using a set of segments; (c) approximation of the shape function associated with each segment by using parametric splines; (d) approximation of the luminosity function associated with each segment by using parametric splines.

In Figure 1, the block diagram of the method is shown where the coding and decoding phases are individuated.

2.1 The coding module
A Change Detection (CD) module is used to compare the input image I(x,y) with a background image S(x,y) to individuate intruder objects (e.g., car, human, etc.) in the surveilled scene [6,7]. The output, given by a binary image X(x,y) where each blobs represents a possible object, is first filtered to reduce the noise [3] and then it is provided as input to a module which extracts the Statistical Morphological Skeleton (SMS), SMS(x,y) [5].
The SMS is obtained through the iterated application of
Binary Statistical Morphology (BMS) operators, i.e.,
Binary Statistical Erosion (BSE) and Binary Statistical
Dilation (BSD) to a shape which is progressively shrank.
The SMS can be described as [4]:

\[
\text{SMS}(x,y) = \{(x,y, n(x,y)) : (x,y) \in \text{I} \}
\]  

(1)

where the function \( n(x,y) \) associates with each skeleton
point the iteration at which the point itself has been
detected.

The SMS of a shape \( X \) is provided as input to a module
whose goal is to efficiently represent information
contained in the SMS itself. This module consists of two
parts: (a) a straight-segment extractor, i.e., the Direct
Hough Transform (DHT) [8], and the (b) a shape
approximation (SA) module. The DHT module takes as
input the position of skeleton points on an image, by
extracting the set \( S = \{ s_j \}_{j=1}^{\text{N}} \) of rectilinear segments, \( s_j \),
which best fits with the \((x,y)\) points of SMS\((x,y)\). The SA
module first attaches the shape information of skeleton to
points of the segments in \( S \), so associating with each
segment \( s_j \) a function \( g_i(\cdot) \) representing the behaviour
of shape-information within the segment itself.

\[
\text{SA}(X) = \{ \{ s_j, g_j(x,y) \} : s_j \in S, (x,y) \in s_j \}
\]  

(2)

Each function \( g_i(\cdot) \) is interpolated by means of a set of
of cubic parametric base-functions (B-spline) [9], whose
number depends on the length of the segment.

A vector \( v_{j,u} \) of \( C_j \) coefficients describes the \( u \)-th spline
of the \( j \)-th segment. The coefficients of the interpolating
functions are associated with each segment to generate a
shape-representation, \( R(X) \):

\[
R(X) = \{ \{ s_j, v_j \} : s_j \in S(X), v_j = [v_{j,u} : u=1,...,C_j] \}
\]  

(3)

Shape representation \( R(X) \) is provided to the
Luminosity Approximation (LA) module which
associates with each segment point \( r \) a vectorial function
\( l_i(r) \) representing the behaviour of the luminosity
function \( f(x,y) \) into a neighbourhood image area \( A_r \) of the \( r \) point
itself:

\[
l_i(r) = l_i(x,y) = u_0 + u_1 x + u_2 y + u_3 x^2 + u_4 x y + u_5 y
\]  

(4)

where the vector of coefficients \( U_j = [u^{(0)},...,u^{(5)}] \) is
obtained by using a polynomial approximation [3] on a
part of the luminosity function. The domain of the
luminosity function is chosen whose shape can be
reconstructed starting from knowledge on the behaviour of
the shape function \( g_i(r) \) in a neighborhood \( A_r \) of the point
\( r \). The order of the neighborhood depends on the
scheduling strategy adopted to extract the SMS. The
components of the vectorial function are approximated
separately by using parametric splines. The LA module
provides as output a representation \( R(L) \) consisting of
both shape and pictorial information. The output is
represented by:

\[
L(X) = \{ \{ s_j, l_i(r) \} : s_j \in S, r \in s_j \}
\]  

(5)

2.2 The decoding module

The decoding module operates by progressively
providing new information depending on the type of
representation considered. First, information on the
object presence and the object type is received. A
transformation \( \Phi \) based on Statistical Morphology
is applied to the received shape representation \( \tilde{R}(X) \) in
order to recover an approximation, \( \tilde{X} \), of the original
binary shape. The \( \Phi \) transformation is composed by the
following steps [3,10]:

(a) \( \text{SMS}(0) = \{(x,y) : (x,y) \in \text{I}, \hat{n}(x,y) = n_{\text{max}} \} \) \( k = 1 \)

(b) \( P(k) = \{(x,y) : (x,y) \in \text{I}, \hat{n}(x,y) = n_{\text{max}} - k \} \)

(c) \( \text{SMS}(k) = ([\text{SMS}(k-1) \ominus B] \cap \text{SMS}(k-1)) \cup P(k) \)

(d) \( k = k+1 \) : if \( k < n_{\text{max}} \) then go to (2)

(5) \( \tilde{X} = \text{SMS}(k-1) \)

where \( B \) is the structuring element [3].

Then, pictorial information can be reconstructed in three
steps.

(a) An estimate of the luminosity function \( \tilde{l}_i(r) \) is
computed for each segment point \( r \) on the basis of
received and estimated coefficients \( \hat{U}_j \).

(b) The behaviour of the luminosity function is
reconstructed within different neighbourhood image areas
\( A_r \) of each segment point \( r \) (i.e., approximated SMS) by
applying a dilation operator \( n(r) \) times [3,10]. The locally
estimated behaviour of the luminosity function is used to
constrain the solution. The constraints, coming from
considering separately each point of the approximated SMS, are used to obtain the global solution.

(c) As some areas $A_r$ can overlap in space, i.e., $A_r \cap A_k \neq 0$, $r,k \in s_p$, a fusion step, consisting of a simple average operator [5.10], is applied to each image point belongs to each $A_r$.

3 RESULTS

A surveillance system for railway level-crossing monitoring has been considered as a real application for testing the proposed method. Fig. 2a shows a 256*256 b/w image of the surveilled level-crossing containing multiple intruder vehicles, while Fig. 2b shows the output of the CD module. Noise generated by fast changing of the lighting conditions and geometric distortions introduced by the TV camera are present [1,2]. The goal of the system is to transmit the shape of the detected intruder object to a remote operator (e.g., people controlling the global railway line in a central control room).

Four significant blobs, representing possible obstacles, are detected. Figure 2c shows the SMSs extracted from each blob, while segments found in S(X) are shown in Figure 2d. A spline characterized by 4 coefficients is used to approximate segments 20 pixels long.

Figs. 2e and Fig. 2f show the reconstructed binary shape $\hat{R}(X)$ and the b/w reconstructed image $\hat{I}$, respectively. Let us observe that the blob c which represents a car is obtained by applying the dilation operator 8, 14 and 7 times to the points of the segments $s_1, s_2, s_3$, respectively.

ACKNOWLEDGEMENTS

This paper has been partially supported by the PROGETTO FINALIZZATO TRASPORTI II (PFT2) of Italian Research Council.

REFERENCES


Fig. 1. Block diagram of the method

Fig. 2 (a) The original image, (b) the change-detection image, (c) the SMSs of the detected objects, (d) the set of segments $S(X)$ approximating the MSs, (e) the reconstructed binary shape $\hat{R}(X)$ and (f) the b/w reconstructed image $\hat{I}$. 