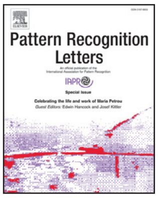
Pattern Recognition Letters 65 (2015) 131–136

Contents lists available at ScienceDirect

Pattern Recognition Letters

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



Highly efficient contour-based predictive shape coding☆

Zhong-jie Zhu[a,]*, Yu-er Wang[a], Gang-yi Jiang[b]

[a] *Ningbo Key Lab of DSP, Zhejiang Wanli University, Ningbo 315100, China*
[b] *Institute of Circuits and Systems, Ningbo University, Ningbo 315211, China*

ARTICLE INFO

Article history:
Received 14 February 2015
Available online 31 July 2015

Keywords:
Image coding
Shape coding
Predictive coding
High efficiency

ABSTRACT

A new scheme for contour-based predictive shape coding is proposed aiming to acquire high coding efficiency, where the temporal correlations among object contours are effectively exploited. For a given binary shape image, the object contours are firstly extracted and thinned to be perfect single-pixel width followed by chain-based representation. Then a chain-based motion estimation and compensation technique is developed to remove temporal correlations among object contours to reduce the data to be encoded. Finally, by further exploiting the spatial correlations within chain links, a novel method is introduced to efficiently encode the residuals together with the motion displacements. Experiments are conducted and the results show that the proposed scheme is considerably more efficient than the existing techniques.

1. Introduction

Shape is an important feature of objects and shape coding is a key technique in object-based image and video applications such as object-based coding, object-based editing and object-based retrieval [1–3]. Since original shape information of objects is usually stored and represented by bi-level images or image sequences, shape coding in essence is a kind of image or video coding, which could be either lossy or lossless. Most of the existing shape coding techniques can be classified into two main categories: bitmap-based methods and contour-based methods. Bitmap-based methods encode a binary shape image directly and they are usually fast and simple and can be used for both lossy and lossless applications. Typical methods of this kind include the JBIG [4], JBIG2 [5,6], MPEG-4 CAE [7,8] and so on. However, their coding efficiency is usually not high. By comparison, contour-based methods only extract and encode the contours of objects instead of encoding the whole shape images, which can usually acquire high coding efficiency. Typical methods of this kind include the chain coding techniques [9] and the curve fitting techniques [10,11]. Generally, the curve fitting techniques can acquire higher compression efficiency than the chain coding techniques. However they are mostly used for lossy coding while the latter can be used for both lossy and lossless coding.

Shape coding is a hot research topic in the field of image and video signal processing. Much work has been done and many schemes have been proposed. In [12], a SPIHT-based scheme for coding both shape and texture of arbitrarily shaped objects was proposed. It employs a novel implementation of the shape-adaptive discrete wavelet transform (SA-DWT) using in-place lifting along with parallel coding of texture coefficients and shape mask pixels aiming to create a single embedded code that allows for fine-grained rate-distortion scalability. Objective and subjective simulation results show it has a rate-distortion performance comparable or superior to MPEG-4 CAE. In [13], an image-dependent shape coding algorithm was proposed which encodes shape as dependent meta data. Both the coding and the decoding processes are designed to be dependent on the underlying image. The correlation between image and shape can be effectively removed and the coding efficiency was greatly improved over most of the existing techniques. However, it is a lossy coding scheme and can't be used for lossless applications.

In [14,15], a context-based lossless coding scheme (DSLSC) for bi-level-images was proposed. It is based on a local analysis of the digital straightness of the causal part of object boundary, which is used in context definition for arithmetic encoding. Experimental results show that it outperforms most of the existing techniques [14,15]. In [16], a block-based scheme that combines a quad-tree structure with context-based arithmetic coding was proposed. Comparing with MPEG-4 CAE, it can considerably improve the coding efficiency. Liu and Ngan explored the H.264/AVC-based shape coding techniques and developed a new scheme that can encode arbitrarily shaped object [17]. Wulandhari and Haron proposed a scheme that can deal with shapes with holes based on vertex chain code [18]. Lai et al addressed the issue of distortion measurement for B-spline-based shape coding and proposed a new metric that is fairly perceptually consistent [19]. In [20], an attempt to perform shape coding using a

---

☆ This paper has been recommended for acceptance by D. Coeurjolly.

* Corresponding author. Tel.: +86 137 770 033-78; fax: +86 574 8822 2542.
E-mail address: zhongjiezhu@hotmail.com (Z.-j. Zhu).

http://dx.doi.org/10.1016/j.patrec.2015.07.026
0167-8655/

multiple grid chain code was proposed, which can support both lossless and quasi-lossless coding.

Due to the importance of shape information to objects, lossless or near lossless shape coding is of vital importance in many practical applications. For example, in pattern recognition applications, accurate shape information is very important to recognize an object. In object-oriented image retrieval applications, the accuracy of shape greatly impacts the retrieval efficiency and accuracy. In medical image applications, accurate shape of an organ can also help to diagnose a disease or undergo an operation. Hence, efficient lossless or near lossless shape coding is desired and worth being further studied. Chain-coding is a popular technique that can be used for lossless shape coding. However, most of the existing chain-based shape coding methods have not exploited the temporal redundancy contained within shape image sequences. Actually, just like the existence of strong temporal redundancy among video textures, there also exist strong redundancies between temporal object contours, which also can be exploited to improve coding efficiency. However, this kind of redundancy is more difficult to be exploited than that within video textures. For video textures, the redundancy among them can be removed and exploited via the well-known block-based inter-frame prediction based on motion-estimation and -compensation techniques. However, due to the difference between video textures and object contours, this kind of techniques can't be directly used for contour-based shape coding.

In this paper, the technique of chain-based predictive shape coding is addressed and a high efficient scheme is proposed for near lossless coding, where the temporal correlations among object contours are efficiently exploited. In our scheme, the object contours of each shape image are firstly extracted and thinned to be perfect single-pixel width followed by chain-based representation. Then a chain-based motion estimation and motion compensation scheme is introduced to efficiently remove the inter-frame correlations to reduce the overall image data to be encoded. Finally, a novel method is developed to efficiently encode the residuals. As a result, the temporal redundancy among object contours can be efficiently exploited and the overall coding efficiency can be considerably improved.

Comparing with the existing shape coding techniques, our main contributions can be summarized as: (1) The technique of contour-based predictive shape coding is addressed and a novel high efficient scheme is proposed. (2) A chain-based motion estimation and motion compensation method is given, which can effectively remove the temporal correlations among object contours. (3) A novel technique is developed for encoding the prediction residuals with high efficiency. The remainder of the paper is organized as follows: Section 2 describes the proposed scheme and details its key steps. Section 3 shows partial experimental results to evaluate the performance of the proposed scheme; Section 4 concludes the paper.

2. Proposed scheme

The diagram of the proposed scheme is shown in Fig. 1, which mainly consists of three major steps: (1) Contour extraction, thinning and chain-based representation; (2) Chain-based motion estimation and compensation; (3) Encoding of residuals and motion displacements.

2.1. Contour extraction, thinning and chain-based representation

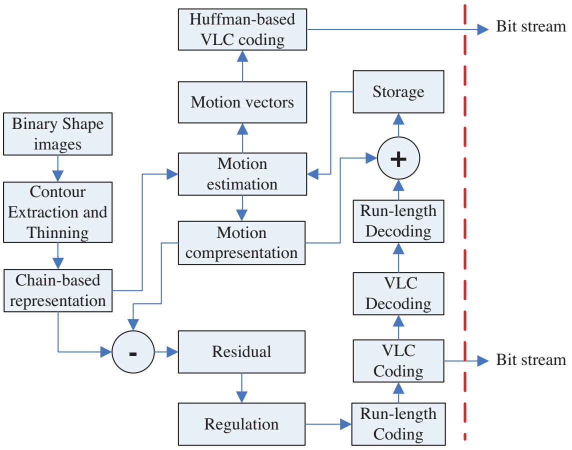


Fig. 1. The diagram of the proposed predictive shape coding scheme.

The first step of chain-based shape coding is to extract the object contours. Many techniques can be employed for contour extraction. In this paper, we use a straightforward one. Let $f(x, y)$ be a given binary shape image and assume the values of foreground object pixels and background pixels are 0 and 255 respectively, the contour extraction is formulated as:

$$C(x,y)=\begin{cases}0 & |f(x,y)|+\left|\prod_{p,q}\left|f(p,q)-255\right|\right|=0\\ 255 & \text{otherwise}\end{cases} \tag{1}$$

where pixels $f(p, q)$ belong to the 4-connected neighborhood of $f(x, y)$ and $C(x, y)$ denote the extracted contour map. If $C(x, y) = 0$, it means pixel $C(x, y)$ is a contour point.

In our scheme, in order to acquire high coding efficiency, the object contours are required to be 8-connected and single-pixel width, that is, there exists only one path between any two neighboring contour points. Hence, the above-extracted contours further performed a thinning operation. Let $C(m, n)$ and$C(p, q)$ be two pixels within the 4-connnected neighborhood of $C(x, y)$, our thinning operation can be formulated as

$$C(x,y)=\begin{cases}255 & \begin{array}{l}|C(x,y)|+|C(p,q)|+|C(m,n)|=0\\ \&\&\sqrt{(p-m)^2+(q-n)^2}=\sqrt{2}\end{array}\\ C(x,y) & \text{otherwise}\end{cases} \tag{2}$$

After thinning operation, the contours are viewed to be perfect single-pixel width. Usually, a point on a perfect single-pixel-width contour has just two neighboring contour points within its 8-connected neighborhood. However, when the contour point is an end-point of an unclosed object contour or is a cross-point of intersectional contours, it may have only one or more than two neighboring contour points. In this paper, for efficient coding, the end-points and the cross-points are selected as key points, based on which the contours are divided into a number of segments to make each segment have zero or only two key points. After thinning and segmentation, each segment is further transformed into chain-based representation based on the well-known 8-connnected Freeman chain code.

2.2. Chain-based motion estimation and compensation

In our scheme, object contours are encoded segment by segment. For a given chain-represented contour segment $L = \{l_n\}(n = 0, \ldots, N-1)$ in the current frame $C_i(x, y)$, where l_n denotes the nth link and N denotes the number of links within L, the steps of its chain-based motion estimation and compensation can be described as follows:

1) Select a segment $L^r = \{l_n^r\}(n = 0, \ldots, N^r - 1)$ for reference from the previously encoded frame $C_{i-1}(x, y)$ based on global matching: