Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 27 (2006) 1252–1265
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Automatic video segmentation using genetic algorithms

Eun Yi Kim [a], Se Hyun Park [b,*]

[a] *Department of Internet and Multimedia Engineering, Next-generation Innovative Technology Research Institute, Konkuk University, 1 Hwayang-dong, Gwangjin-gu, Seoul, Republic of Korea*
[b] *School of Computer and Communication, Daegu University, Republic of Korea*

Available online 5 January 2006

Abstract

The current paper proposes a genetic algorithm (GA)-based segmentation method that can automatically extract and track moving objects. The proposed method mainly consists of spatial and temporal segmentation; the spatial segmentation divides each frame into regions with accurate boundaries, and the temporal segmentation divides each frame into background and foreground areas. The spatial segmentation is performed using individuals that evolve distributed genetic algorithms (DGAs). However, unlike standard DGAs, the individuals are initiated from the segmentation result of the previous frame, then only unstable individuals corresponding to actual moving object parts are evolved by mating operators. For the temporal segmentation, adaptive thresholding is performed based on the intensity difference between two consecutive frames. The spatial and temporal segmentation results are then combined for object extraction, and tracking is performed using the natural correspondence established by the proposed spatial segmentation method.

The main advantages of the proposed method are twofold: first, proposed video segmentation method does not require any a priori information; second, the proposed GA-based segmentation method enhances the search efficiency and incorporates a tracking algorithm within its own architecture. These advantages were confirmed by experiments where the proposed method was successfully applied to well-known and natural video sequences.

Keywords: Genetic algorithm; Markov random fields; Object detection; Tracking

1. Introduction

Video segmentation has been the subject of intensive research due to its importance in a variety of applications, including vision systems, pattern recognition, and the MPEG-4 coding standard (Sikora, 1997; ISO/IEC JTC1/SC29/WG11, 1998; Salembier and Marques, 1999; Kim et al., 1999; Gu and Lee, 1998; Bovik, 2000).

So far many techniques have been proposed for video segmentation, and proved to be very efficient in a supervised mode (Salembier and Marques, 1999; Kim et al., 1999; Gu and Lee, 1998; Bovik, 2000). However, automatic video segmentation is still a difficult and unresolved problem. The crucial issues related to automatic video segmentation are (1) separating moving objects from the background and (2) obtaining accurate objects boundaries. To satisfy thee requirements, both the motion to identify objects in a scene and the colors to provide important hints about object boundaries should be used (Salembier and Marques, 1999; Kim et al., 2000b, 2001; Mocheni et al., 1998; Meier, 1998). As such, spatiotemporal segmentation that uses both color and motion has become a recent focus of intensive research and would seem to produce robust segmentation results (ISO/IEC JTC1/SC29/WG11, 1998; Salembier and Marques, 1999; Kim et al., 1999, 2001; Mocheni et al., 1998; Meier, 1998). Generally it consists of three major modules: spatial segmentation, temporal segmentation, and the fusion of spatiotemporal segmentation and temporal coherence. Spatial segmentation divides each frame into regions with accurate boundaries, while temporal segmentation labels the pixels in each frame that are associated with independently moving parts of the scene. Objects forming each frame are then created by

* Corresponding author. Tel./fax: +82 53 850 6637.
E-mail address: sehyun@daegu.ac.kr (S.H. Park).

0167-8655/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2005.07.023

combining the spatial and temporal segmentation result, thereafter tracked based on the temporal coherence of the objects between successive frames.

Among the many spatial segmentation methods, morphological and Bayesian approaches have gained most popularity over the last few years (Salembier and Marques, 1999; Bovik, 2000; Meier, 1998). In morphological approaches, the initial regions are first selected, then a watershed algorithm is used to label the pixels (Salembier and Marques, 1999; Meier, 1998). The approaches are computationally efficient. However, they suffer from over-segmentation, thereby requiring additional region-merging techniques (Salembier and Marques, 1999; Bovik, 2000; Meier, 1998). Another disadvantage is the lack of constraints to enforce spatial continuity in the segmentation. Bayesian approaches perform a maximum a posteriori (MAP) estimation of unknown segmentation label fields, given the observed video (Kim et al., 2000a,b, 2001; Li, 1995; Andrey and Tarroux, 2000; Pal and Wang, 1996). Accordingly, the segmentation is formulated as an optimization problem and a solution is obtained using optimization algorithms. A clear advantage of Bayesian approaches over morphological approaches is the incorporation of spatiotemporal continuity constraints (Bovik, 2000; Meier, 1998). Yet, intensive computational costs are the main shortcomings. Therefore, the major issue for this type of method is the selection of a good optimization algorithm. Recently, genetic algorithms (GAs) have appeared as a good solution because they are robust and can successfully deal with combinatorial problems (Kim et al., 2000a; Andrey and Tarroux, 2000; Pal and Wang, 1996; Park et al., 1997; Bhandarkar and Zhang, 1999). GAs are also attractive as they can achieve an efficient parallel exploration of the search space without being confined to local optima. As such, GAs have already been used in many segmentation problems (Kim et al., 2000a; Andrey and Tarroux, 2000; Pal and Wang, 1996; Park et al., 1997; Bhandarkar and Zhang, 1999).

There are also many possible choices for temporal segmentation. The simplest and most widely used approach is comparing the temporal differences of the intensity values between two consecutive frames, (ISO/IEC JTC1/SC29/WG11, 1998; Kim et al., 1999; Bovik, 2000; Meier, 1998). In practice, this difference is dependent on the activity of the objects in the scene. Therefore, if an object moves very slowly or stops at a given time, it can be mistaken for background. In a generic video sequence, a moving object will often move continuously through a whole video sequence. Accordingly, the missing problem can be easily solved based on this region-tracking information.

This paper presents a new form of spatiotemporal segmentation that can automatically decompose a video into the objects making up the scene. For the spatial segmentation, a GA-based method is proposed that includes new evolution mechanisms to improve the performance. The segmentation is performed by individuals that independently evolve using distributed genetic algorithms (DGAs). However, unlike conventional DGAs, the individuals are initialized using the segmentation results of the previous frame, then only unstable individuals corresponding to actually moving object parts are evolved by crossover and mutation. These mechanisms facilitate the effective exploitation and exploration of the search space. They also enables those regions corresponding to the same objects in successive frames to retain the same labels through a whole video sequence, which is used as important cues for the proposed object-tracking method. For the temporal segmentation, coherent segmentation results are produced by first adaptive thresholding the intensity difference between two consecutive frames, then the results are verified and compensated by incorporating historical information where each pixel is labeled as changed if it belonged to an object in the previous frame. Finally, objects are extracted by combining the spatial and temporal segmentation results, and then tracked by the natural correspondence established in the spatial segmentation. The proposed method is able to deal with scenes that include multiple objects, and keep track of objects even when they stop moving for an arbitrarily long time.

The organization of the paper is as follows. Section 2 describes the proposed spatial segmentation method, while the temporal segmentation is presented in Section 3. The extraction and tracking of objects are included in Sections 4 and 5. Section 6 presents the experiments followed by a conclusion.

2. Spatial segmentation

2.1. Image modeling using spatiotemporal MRFs

Let S be the 3-D volume of $M_1 \times M_2$ lattices, such that an element S_{st} indexes a pixel at site s and time t. Let G be the input sequence of color images defined on S. Here, the input image G was considered as degraded by i.i.d. (independent identically distributed) zero-mean Gaussian white noise $N = \{n_{st}\}$. Then, $X = \{X_{st} \mid X_{st} \in \Lambda \text{ and } 1 \leqslant s \leqslant M_1 \times M_2\}$ denotes the label configurations defined on S, wherein X_{st} is a discrete random variable taking a value in the label set $\Lambda = \{\lambda_1, \ldots, \lambda_R\}$. A spatiotemporal neighborhood in S is defined as $\Gamma = \{\eta_{st}\}$, where η_{st} is the set of neighboring sites (s,t). Then, X is modeled by a spatiotemporal MRF based on S with respect to Γ, because it satisfies the following spatiotemporal Markovian property (Murray and Buxton, 1990).

$$P(X_{st} = x_{st} | X_{qr} = x_{qr}, (s,t) \neq (q,r)) \\ = P(X_{st} = x_{st} | X_{qr} = x_{qr}, (q,r) \in \eta_{st}).$$

Let ω be a realization of X. Then, the segmentation is to estimate the label configuration ω that maximizes the following posterior distribution for a fixed input video g.

$$\arg\max_{\omega} P(X = \omega | G = g) = \arg\max_{\omega} \frac{P(g|\omega)P(\omega)}{P(g)}. \tag{1}$$