PERGAMON

Pattern Recognition 34 (2001) 2063–2066

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Rapid and Brief Communication

Spatiotemporal segmentation using genetic algorithms

Eun Yi Kim[a], Sang Won Hwang[a], Se Hyun Park[b], Hang Joon Kim[a,*]

[a]*Department of Computer Engineering, Kyungpook National University 1370, Sangyuk-dong, Puk-gu Taegu 702-701, South Korea*
[b]*Division of Computer Engineering, Chosun University, 375 Seosuk-dong, Dong-gu, Kwangju, South Korea*

Received 19 July 2000; accepted 10 August 2000

1. Introduction

Segmentation is the process of identifying uniform regions based on certain conditions. Segmentation has been used for a long time in image analysis and computer vision for a variety of applications. In particular, there has been a growing interest in video sequence segmentation mainly due to the development of MPEG-4, which enables the content-based manipulation of multimedia data [1,2]. For this, the sequence must be first segmented into a set of meaningful objects [1,2]. However, since such objects are normally not characterized by a homogeneous intensity, color, or optical flow, conventional segmentation algorithms using these features cannot produce meaningful partitions [1,2]. To overcome this problem, motion information has been used recently in many segmentation techniques, as one of the most important characteristics for identifying objects in a scene [1]. The methods that use motion can be divided into two main classes: joint motion estimation and segmentation, and spatiotemporal segmentation. The first approach estimates motion vectors at the pixel level and then clusters the pixels into regions of coherent motion, however, it has a major drawback: regions with coherent motion may contain multiple objects and need further segmentation for object extraction [1]. To overcome this drawback, the second approach uses information from both the spatial and temporal domain (i.e. color and motion), which produces a more meaningful segmentation [1,2].

This paper presents a new spatiotemporal segmentation of a video sequence. Each frame in a sequence is modeled using a Markov random field (MRF), which is effective in describing the spatial and temporal dependency of neighboring pixels and robust to degradation. Since this is computationally intensive, a new genetic algorithm (GA)-based segmentation method that can improve the computational efficiency is developed.

In the proposed method, the energy function of each MRF is minimized by chromosomes that evolve using distributed genetic algorithms (DGAs). Then, to eliminate any redundant computation among consecutive frames, only unstable chromosomes corresponding to moving object parts are evolved. Experimental results show that the proposed method can improve the computational efficiency as well as the quality of the segmentation result.

2. Image modeling using MRF

The input image G was considered as degraded by i.i.d. (independent identically distributed) zero-mean Gaussian white noise $N = \{n_{ij}\}$. Let $S = \{(i, j): 1 \leqslant i \leqslant M_1, 1 \leqslant j \leqslant M_2\}$ denote the $M_1 \times M_2$ lattice, such that an element in S indexes an image pixel. Let $\Lambda = \{\lambda_1, \ldots, \lambda_R\}$ denote the label set and $X = \{X_{ij} | X_{ij} \in \Lambda\}$ be the family of random variables defined on S. The neighborhood of S can be defined as $\Gamma = \{\eta_{ij}\}$, where η_{ij} is a set of sites neighboring (i, j). Then, X is an MRF on S with respect to Γ because the two conditions of [3] are satisfied, Let ω be a realization of X. The goal is to identify ω which maximizes the posterior distribution for a fixed input image g. That is, to determine

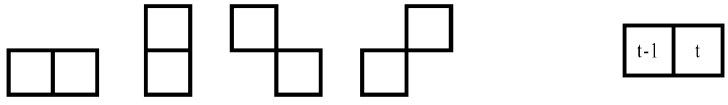
$$\arg\max_{\omega} P(X = \omega | G = g) \approx \arg\max_{\omega} P(g|\omega)P(\omega). \tag{1}$$

Eq. (1) is divided into two components follows:

$$\begin{aligned} \mathrm{P}(g|\omega) &= P(N = g - F(\omega)|\omega) \\ &= \prod_{(i,j)\in S} P(n_{ij} = g_{ij} - F(\omega_{ij})|\omega_{ij}) \\ &= \prod_{(i,j)\in S} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(g_{ij} - F(\omega_{ij}))^2}{2\sigma^2} \right], \end{aligned} \tag{2}$$

* Corresponding author. Tel.: + 82-53-950-5554; fax: + 82-53-957-4846.
E-mail address: kimhj@ailab.knu.ac.kr (H.J. Kim).

PII: S0031-3203(00)00129-1



(a) (b)

Fig. 1. Cliques: (a) spatial cliques; (b) temporal clique.

$$P(\omega) = \exp(-U(\omega)) = \exp\left\{-\sum_{c\in C}[S_c(\omega) + T_c(\omega)]\right\}. \quad (3)$$

In Eq. (2), σ is the noise variance and $F(\cdot)$ is the mapping function that the label of a pixel corresponds to the estimated color vector. In Eq. (3), C is a possible set of cliques, where a clique is defined as a set of pixels in which all the pairs are mutual neighbors. Then the energy function $U(\omega)$ is obtained by the summation of two potentials over all possible cliques: spatial potentials $S_c(\omega)$ and temporal potentials $T_c(\omega)$. The former imposes the spatial continuity of the labels and the latter is to achieve the temporal continuity of the labels. The proposed model assumes that the only nonzero potential are those corresponding to two-pair cliques, as shown in Fig. 1. Then $S_c(\omega) = -\alpha$ if all labels in c are equal, otherwise $S_c(\omega) = \alpha$. Similarly, $T_c(\omega) = -\beta$ if all labels in c are equal, otherwise $T_c(\omega) = \beta$.

Accordingly Eq. (1) can be represented by the following equation, which is defined as a posterior energy function:

$$\arg\min_{\omega}\left\{\sum_{c\in C}[S_c(\omega) + T_c(\omega)] + \frac{[g - F(\omega)]^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2)\right\}. \quad (4)$$

Let ρ_{ij} denote a set of cliques containing pixel (i, j). Since C is equal to the sum of ρ_{ij} for all pixels, the function in Eq. (4) can be rewritten as the sum of the local energy U_{ij} for all pixels:

$$\arg\min_{\omega}\sum_{(i,j)\in S}\left\{\sum_{C\in\rho_{ij}}[S_c(\omega_{ij}) + T_c(\omega_{ij})] + \frac{[g_{ij} - F(\omega_{ij})]^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2)\right\}. \quad (5)$$

As a result, instead of maximizing the posterior distribution, the posterior energy function is minimized to identify the optimal label.

3. Proposed segmentation algorithm

In this paper, the segmentation problem of a frame is formalized as an optimization problem of the posterior energy function. For this, GAs are used. The computation is distributed into chromosomes that evolve by DGAs. A chromosome consists of a label and a feature vector, which are described in [4]. Its fitness is defined as the local energy U_{ij}. A set of chromosomes is called a population and represents a segmentation result. According to their fitnesses, the chromosomes are classified into two groups: stable and unstable chromosomes. When the chromosomes are mapped to an actual video sequence, the stable and unstable chromosomes correspond to the background and moving object parts, respectively.

In the proposed method, the segmentation of the frames in a sequence is successively obtained. For the first frame, the chromosomes are initiated with random values, whereas, for later frames they are started from the segmentation of the previous frame. The segmentation for the starting frame is determined using the Kim et al. segmentation algorithm [4]. Thereafter, the remaining frames are segmented using the algorithm outlined as in Fig. 2.

At intervals of T, the system receives the input $I(t)$ and $C(t-l)$, that is, the current frame and segmentation result of the previous frame, respectively. Starting with the segmentation results of the previous frame, the chromosomes are evolved through iteratively performed selection and genetic operations. In DGA, these operations are performed on locally distributed subgroups of chromosomes, called windows, rather than on whole populations. In the selection process, the chromosomes are updated to new chromosomes. $C'(t)$, by an elitist selection scheme [4]. Thereafter, in the Decision Module, the

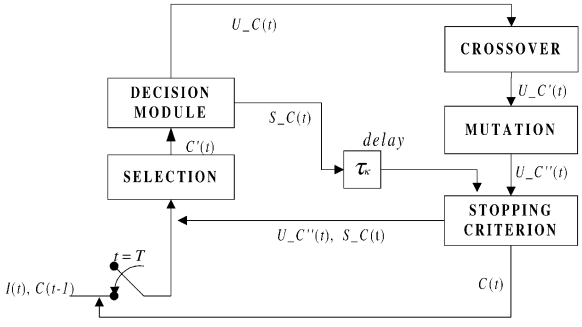


Fig. 2. Block diagram of proposed algorithm.

chromosomes are classified into two groups according to their fitness: stable chromosomes, $S\_C(t)$, and unstable ones, $U\_C(t)$. In the current frame, the chromosomes are stored in an increasing order based on their fitnesses. Given the probabilities of genetic operations, certain chromosomes with lower fitnesses are selected as being unstable: prior to crossover and mutation, the unstable chromosomes are determined in accordance with their probabilities. Only the unstable chromosomes are evolved by crossover and mutation, which are described in detail in [4]. It should be noted that the probabilities of genetic operations have an influence on the performance of the algorithms. Accordingly, these parameters were determined empirically. In Fig. 2, $U\_C'(t)$ and $U\_C''(t)$ are chromosomes evolved by crossover and mutation, respectively. $S\_C(t)$ is then delayed for τ_k, the time taken for the genetic operations within a generation. These operations are iteratively performed until the stopping criterion is satisfied. For the stopping criterion, the equilibrium is defined in [4]. The stopping criterion is reached when the equilibrium is above the equilibrium threshold or the number of generations is more than the maximal number.

4. Results

To assess the validity of the proposed method, it was tested on several well-known video sequences. The experiments were performed on a Pentium II-233 PC. The equilibrium threshold was set to 99.5% and the maximal number of generations was set to 1000. In the proposed model, α, β and σ are significant parameters. The value of α has an affect on the number of regions in a frame, and β encourages pixels to have the same label in consecutive frames. Therefore, α and β were empirically determined. The noise variance, σ was obtained as follows: let σ_{ij} be the color variance of the pixels inside the window centered at (i, j) and σ_m be an average of σ_{ij}; then σ is defined as an average of $\{\sigma_{ij} | \sigma_{ij} < \sigma_m\}$.

Fig. 3 shows the segmentation results of "Table Tennis". All frames in the sequence were color images sized at 351×239. The parameters were as follows: α was 1.2; β was 1; the window size was 5×5; the crossover and mutation probabilities were 0.1 and 0.01, respectively. Plus $|\Lambda| = 64$. Then, Fig. 3b shows segmentation results of the proposed method. Fig. 3b shows the frames were perfectly segmented into six objects: the background, arm, ball, racquet, right hand, and left hand. Moreover, it can be found that the results correctly tracked ball's motion and hand's motion even though only the unstable chromosomes were evolved.

The proposed method was compared with the Kim et al. method relative to speed and quality [4]. They proposed an MRF-model-based segmentation method using GAs, which used only color information and independently segmented each frame in a sequence using

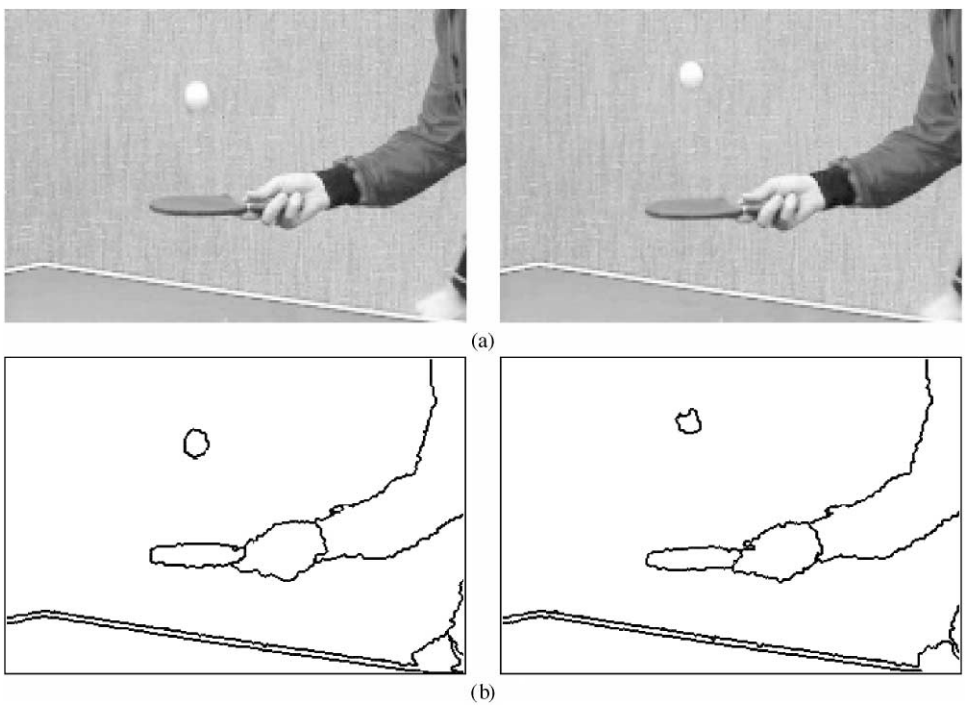


Fig. 3. Segmentation results: (a) input frames; (b) segmentations.

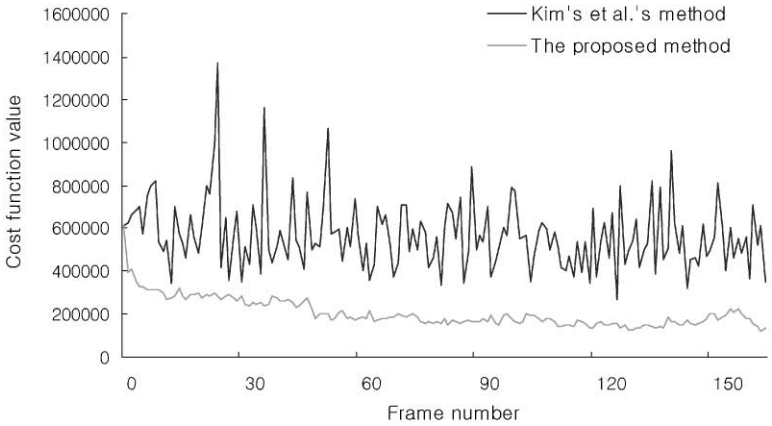


Fig. 4. Quantitative comparison between two methods using cost function.

Table 1
Time taken to segment a frames [s]

	Method	Kim et al.'s method	Proposed method
Average time for generation	Evaluation	0.26	0.26
	Selection	0.05	0.04
	Crossover	0.02	0.01
	Mutation	0.01	0.002
Average number of generations		161	10
Average time for frame		54.74	3.12

Table 2
Evaluation of segmentation results

Method	Kim et al.'s method	Proposed method
Average value of F	34.01	23.54

GAs. Table 1 shows the average generation time and time taken to segment a frame for the two methods.

To evaluate the quality of the segmentations, quantitative comparisons were performed based on a uniform evaluation function and cost function minimization. The evaluation function F was proposed by Liu et al., which was also used in [3,4]. Table 2 shows the average values of F for the segmentation results of "Table Tennis". The smaller value of F; the better the segmentation results [3]. Fig. 4 shows the comparison of the final cost function values for each frame in the sequence. In this paper, a posterior energy function was used as the cost function. On average, the proposed method achieved a cost function that was 63% lower than that obtained by the Kim et al. method.

Fig. 4, Tables 1 and 2 shows that the proposed method not only reduced the computational time, but also improved the quality of the segmentation results.

References

[1] T. Meier, K.N. Ngan, Automatic segmentation of moving objects for video object plane generation, IEEE Trans. Circuit Syst. Video Technol. 8 (5) (1998) 525–538.

[2] Gene K. Wu, T.R. Reed, Image sequence processing using spatiotemporal segmenation, IEEE Trans. Circuit Syst. Video Technol. 9 (5) (1999) 798–807.

[3] Jianqing Liu, Yee-Hong Yang, Multiresolution color image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 16 (7) (1994) 689–700.

[4] H.J. Kim, E.Y. Kim, J.W. Kim, S.H. Park, MRF model based image segmentation using hierarchical distributed genetic algorithm, IEE Electronics Letters 34 (25) (1998) 1394–1395.