A NOVEL APPROACH OF FUSING TEMPORAL AND SPATIAL INFORMATION FOR SEGMENTING MOVING OBJECTS *

LIU Tian-Ming QI Fei-Hu ZHAN Yi-Qiang
(Department of Computer Science & Engineering, Shanghai Jiaotong University, Shanghai, 200030, China)

Abstract A novel approach based on spatial-temporal information fusion is proposed to segment moving objects in video sequences. In the proposed approach, temporal and spatial information are integrated gradually by a region binding process during the segmentation, which is different to the process of combining temporal and spatial segmentation results. The significance of the proposed region binding algorithm is that regions are represented and characterized distributively. In the first stage of the proposed approach, current frame is segmented into many small regions primitively. Then, these small regions are bound to form the Binding-CCores (BC). Finally, the rest regions are bound to their neighboring BCs under strong or weak rules to get the final object regions. Experimental result demonstrated the performance of the approach.

Key words moving object segmentation, region binding, HOS test, information fusion.

一种融合时域和空域信息的运动目标分割新方法 *

刘天明 戚飞虎 詹翊强
(上海交通大学计算机科学与工程系,上海,200030)

摘要 提出了一种融合时域和空域信息的方法,用于从视频序列中分割出运动物体. 该方法是在分割过程中通过区域捆绑逐步融合时域和空域信息,而不是在时域分割结束之后再融合空域信息. 分布式地表达分割物体并刻画其特征是区域捆绑的主要特征. 本文的方法首先通过早期分割得到许多小区域,然后将这些小区域捆绑成一些捆绑核,再将剩下的区域通过强或弱的规则捆绑到相邻的捆绑核,从而实现目标区域的分割. 实验结果显示了该方法的良好性能.

关键词 运动物体分割,区域捆绑,HOS 测试,信息融合.

Introduction

Temporal segmentation methods can label stationary and moving areas, but they often fail to provide accurate boundaries. Spatial segmentation usually supplies reliable boundaries, while it suffers greatly from over-segmentation. To automatically segment moving objects in video sequences, several approaches have been proposed to combine temporal segmentation and spatial segmentation results[1]. The main idea of these approaches is the fusion of various intermediate results. However, temporal and spatial information are not fused until intermediate results are available in those approaches. In this paper, the authors propose a novel approach that combines temporal and spatial information during the whole segmentation process.

Region-based segmentation methods are widely used for spatial segmentation. They rely on the postu-

* The project supported by The National Natural Science Foundation of China(No. 60072029)
Received 2000 - 05 - 15, revised 2002 - 12 - 30

* 国家自然科学基金(批准号 60072029)资助项目
稿件收到日期 2002 - 05 - 15, 修改稿收到日期 2002 - 12 - 30

late that neighboring pixels within one region have similar intensity. Split-and-merge[3~5], region merging[6~8,11,12], and region growing[2,9,10] are among the region-techniques. A common ground of these region merging and growing techniques is that the characteristic of the new merged union of each step and the new homogeneity measure between the new union and its neighbors are recalculated. This results in expensive computation cost. In addition, the region characteristic calculated through a particular region model becomes less reliable as it tends to be larger. In this paper, we proposed a novel technique: Region Binding. These regions are distributedly represented and characterized, which distinguishes the region binding from the region merging and region growing.

1 Region binding technique

1.1 Introduction

In the region binding, any bound region union is distributedly represented by its primitive sub-regions and the homogeneity measure of neighboring regions is based on the directly adjacent primitive sub-regions. A bound region union is called Compound-Region (CR) and a primitive sub-region is called Unit-Region (UR) here. By fusing both temporal and spatial information, primitively segmented regions are bound to form the Binding-Cores (BC), whose role is similar to that of seeds in the region growing. Then the rest regions are bound to their neighboring BCs under strong or weak rules. The approach is composed of four stages.

1.2 Representation of CR and homogeneity measure definition

It is assumed that the set of primitive URs is $\{R_M^1, R_M^2, \cdots, R_M^M\}$ and the current segmentation map is $\{R_{M_*}^1, R_{M_*}^2, \cdots, R_{M_*}^{M_*}\}(M_* < M)$. A CR $R_{M_*}^k$ is distributedly represented by its URs, denoted by $\{R_M^{u1}, R_M^{u2}, \cdots, R_M^{u_i}\}$. So we have

$$R_{M_*}^k = \bigcup_{u_*=1}^{u_i} R_M^{u_*} \text{ and } R_M^{u_n} \cap R_M^{u_m} = \phi, \text{ for } \forall n, m \in \{1,2,\cdots i\} \text{ and } n \neq m \tag{1}$$

In region merging or growing techniques, a region model $\mu(\cdot)$ is usually used to characterize the region. For example, $\mu(\cdot)$ may be the intensity average of a region[8]. In the proposed region binding technique, the region model is used only for URs. A CR does not have a region model and is distributedly characterized by the model of its URs. The distributed representation of CRs results in distributed homogeneity measures. For a pair of neighboring region $R_{M_*}^k$ and $R_{M_*}^p$, there are three modes of homogeneity measure definition. Mode 1: if $R_{M_*}^k$ and $R_{M_*}^p$ are both URs, the homogeneity measure is defined by them directly. Mode 2: if one of them, for example $R_{M_*}^k$, is a CR, the homogeneity measure is defined by one UR of $R_{M_*}^k$ and $R_{M_*}^p$. For $R_{M_*}^p$, there exists at least one UR of $R_{M_*}^k$ that is directly adjacent to it. Find the UR R_M^{uo} that minimizes the model difference between R_M^{uo} and $R_{M_*}^p$. The uo satisfies

$$uo = \arg\{\min_{n=0,1,\cdots,u_i} \| \mu(R_M^p) - \mu(R_M^n) \| \}, \tag{2}$$

Then, the homogeneity measure between $R_{M_*}^k$ and $R_{M_*}^p$ is defined by R_M^{uo} and $R_{M_*}^p$. Mode 3: if both $R_{M_*}^k$ and $R_{M_*}^p$ are CRs, the homogeneity measure is defined by their directly adjacent UR pair that has the smallest model difference. Let $\{R_M^{v1}, R_M^{v2}, \cdots, R_M^{vj}\}$ be all of the URs that compose $R_{M_*}^p$. Find the pair of adjacent URs, denoted by R_M^{v*} and R_M^{uo}, that has the smallest model difference among all adjacent pairs.

$$(v*, uo) = \arg\{\min_{(v*,uo) \in \{v1,v2,\cdots,vj\} \times \{u1,u2,\cdots ui\}} \| \mu(R_M^{v*}) - \mu(R_M^{uo}) \| \}, \tag{3}$$

where R_M^{uo} and R_M^{v*} are adjacent. The homogeneity measure between $R_{M_*}^k$ and $R_{M_*}^p$ is defined by R_M^{uo} and R_M^{v*}. Then, the binding process may be in a region merging fashion or in a region growing fashion. The binding process may be RAG based[8], NNG based[8], RSST based[12] or FRSST based[11].

2 Fusion of temporal and spatial information

The whole information fusion process is composed of four stages. By fusing both temporal and spatial information, primitively segmented regious are bound to form the Binding-Cores (BC), whose role is similar to that of seeds in the region growing. Then the rest regions are bound to their neighboring BCs under strong or weak rules.

2.1 Stage 1

The input image $f^{(1)}(x,y)$ is morphologically filtered by the operation of opening-closing by reconstruction: $\gamma^{(rec)}(\varepsilon_n(f^{(1)}),f^{(1)})$ and $\varphi^{(rec)}(\delta_n(f^{(1)}),f^{(1)})$, which simplifies the image without corrupting the contour information[13]. The operations of opening-closing by reconstruction are as follows.

$$\gamma^{(rec)}(f,r) = \delta^{(\infty)}(f,r) = \cdots\delta^{(1)}(\cdots\delta^{(1)}(f,r)\cdots,r), \tag{4}$$

and

$$\varphi^{(rec)}(f,r) = \varepsilon^{(\infty)}(f,r) = \cdots\varepsilon^{(1)}(\cdots\varepsilon^{(1)}(f,r)\cdots,r), \tag{5}$$

The output $f^{(2)}(x,y)$ is further simplified by a non-linear function S:

$$\begin{aligned} f^{(3)} &= S(f^{(2)}) \\ &= \begin{cases} f^{(2)} - f^{(2)} \bmod D & if\, f^{(2)} \bmod D < D/2 \\ f^{(2)} + D - f^{(2)} \bmod D & otherwise \end{cases}, \end{aligned} \tag{6}$$

The function S significantly simplifies $f^{(2)}(x,y)$ while reserving the contour information well. Then each flat region, called Unit-Region (UR) here, is labeled with a fast flat region labeling algorithm. This method has a good performance in obtaining the set of primitive regions $\{R_m^1,\ R_m^2,\ \cdots,\ R_m^m\}$ under the assumption that the image is not highly textured and demands low computation. Each UR is spatially characterized by its intensity average $\mu(R_m^k)\,(k \leqslant m)$ computed on the original image $f^{(1)}(x,y)$.

In the following step, rough moving object mask is obtained by the HOS (High Order Statistics) test performed on the inter-frame differences[14]. The difference image between two successive frames is modeled as

$$d_k(x,y) = f_k(x,y) - f_{k-1}(x,y), \tag{7}$$

The changes produced by object movements appear in the inter-frame differences as highly structured components, whose statistical behavior strongly deviates from Gaussianity. The problem of detecting the inter-frame variations produced by moving objects is an instance of the detection of a partially modeled stochastic non-Gaussian signal in Gaussian noise[14]. Firstly, for each pixel (x,y) the fourth-order moment of each inter-frame difference $d(x,y)$ is estimated on a moving 3×3 window $\eta(x,y)\,(N_\eta = 9)$.

$$\hat{m}_d^4(x,y) = \frac{1}{N_\eta}\sum_{(s,k)\in\eta(x,y)}(d(s,k) - \hat{m}_d)^4, \tag{8}$$

where $\hat{m}_d(x,y)$ is the sample mean of $d(x,y)$.

$$\hat{m}_d(x,y) = \frac{1}{N_\eta}\sum_{(s,k)\in\eta(x,y)} d(s,k), \tag{9}$$

Then, the fourth-order moment is compared with a threshold pixel by pixel. The threshold is proportional to the square of the noise variance $\hat{\sigma}_{0d}^2$.

$$\begin{aligned} &H_1 : \hat{m}_d^4(x,y) > c(\hat{\sigma}_{0d}^2)^2 \\ &H_0 : \hat{m}_d^4(x,y) < c(\hat{\sigma}_{0d}^2)^2, \end{aligned} \tag{10}$$

Here, H_0 denotes the hypothesis associated to the "still background" class and H_1 the hypothesis associated to "foreground or covered/discovered background" calss. The constant c is approximately independent from the sequence characteristics and its value has been optimized in the basis of a statistical analysis performed during the experimental activity[15]. To adapt automatically the detection scheme to different kinds of video sequences, the noise variance $\hat{\sigma}_{0d}^2$ in above equation is estimated on a subset S' of the static background.

$$\hat{\sigma}_{0d}^2 = \frac{1}{N_{S'}}\sum_{(s,k)\in S'}(d(s,k) - \hat{m}_d)^2. \tag{11}$$

The details of the choice of S' are referred to[14]. Then the percentage of pixels detected as foreground by the HOS test of each UR, denoted by $FPP(P_m^k)$, is calculated. So each UR is temporally characterized by the foreground pixel percentage (FPP).

2.2 Stage 2

In this stage, temporal and spatial information is fused to form the Binding-Cores (BC) by binding conterminous regions under strong rules in a region merging fashion. Firstly, neighboring regions with similar intensity average and the same FPP are bound. If two neighboring regions R_{m*}^k satisfy,

$$\begin{aligned} &\mathrm{FPP}(P_{m*}^k) = \mathrm{FPP}(R_{m*}^p) \\ &\qquad = 0 \textit{ or } 1 \textit{ and } \mathrm{SPD}(R_{m*}^k, R_{m*}^p) < \lambda_1, \end{aligned} \tag{12}$$

they are bound. SPD(·) is the mean intensity difference of two regions and λ_1 is the intensity threshold. The FPPs of newly bound CRs are recursively computed from their URs. Any pair of conterminous regions are bound if they satisfy the above rule. Secondly, any pair of neighboring regions that are both larger than a threshold are bound if they and all of their neighboring regions are temporally homogeneous. In this step, temporal homogeneity is emphasized because the spatial homogeneity of a moving object or background in a

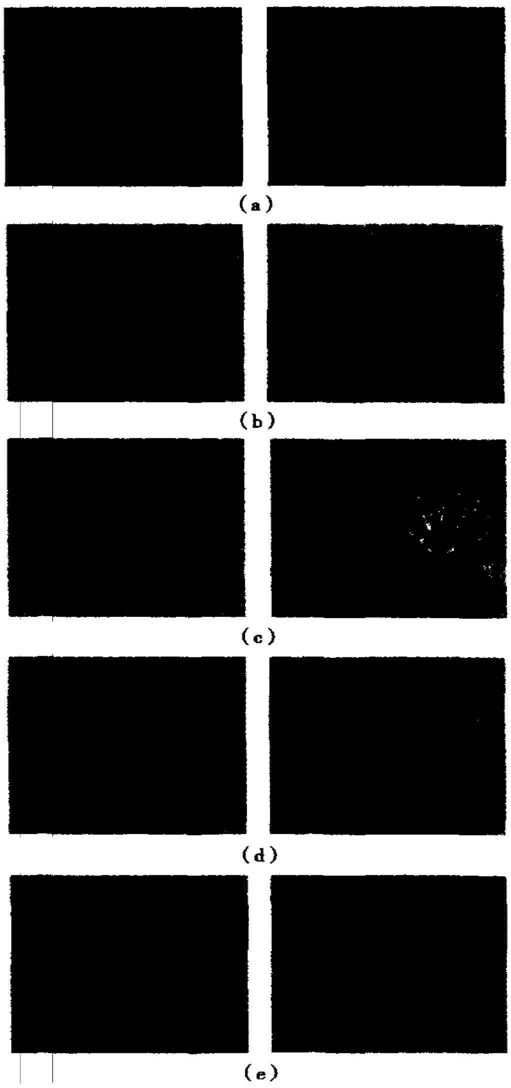

Fig. 1 The segmentation results of akiyo and mother & Daughter
(a) original frames
(b) segmented objects by the HOS test
(c) BCs of foreground after stage 2
(d) BCs of foreground after stage 3
(e) final segmented objects
图1 Akiyo 和 mother & daughter 的分割结果
(a)原始帧
(b)HOS 测试后分割的物体
(c)第二步后前景的 BC
(d)第三步后前景的 BC
(e)最后的分割结果

global sense does not make any sense. Contextual temporal information is also considered to avoid binding foreground and background regions as the segmented object masks by the HOS test are slightly too large[14]. Large regions produced in this stage are labeled as BCs. A fundamental and challenging problem is to select suitable seeds in region growing techniques[9]. Here, the process of forming BCs is automated by combining temporal and spatial information.

2.3 Stage 3

Strong rules are used to bind regions inside foreground or background in this stage. Firstly, all of the regions that are both temporally and spatially homogeneous with their neighboring BCs are bound to them respectively in a region growing fashion. For a given BC R^{k}_{m*}, whose neighbors are $\{R^{k1}_{m*}, R^{k2}_{m*}, \cdots, R^{kj}_{m*}\}$, if any of its neighbors R^{ki}_{m*} $(i \leqslant j)$ satisfies

$$|\mathrm{FPP}(R^{k}_{m*}) - \mathrm{FPP}(R^{ki}_{m*})| < \lambda_2 \text{ and } \mathrm{SPD}(R^{k}_{m*}, R^{ki}_{m*}) < \lambda_3, \qquad (13)$$

R^{ki}_{m*} is bound to R^{k}_{m*}. λ_2 is the FPP threshold and λ_3 is the intensity threshold. At the same time, those regions that are both temporally and spatially homogeneous with internal to BCs are bound to them. Secondly, any pair of neighboring BCs with temporal homogeneity are bound to form larger binding-cores. The above two steps are iterated until no qualified pair of regions exists.

2.4 Stage 4

Firstly, small CRs are unbound because using intensity averages of URs is more reliable and convenient. Then a procedure similar to stage 3 is performed

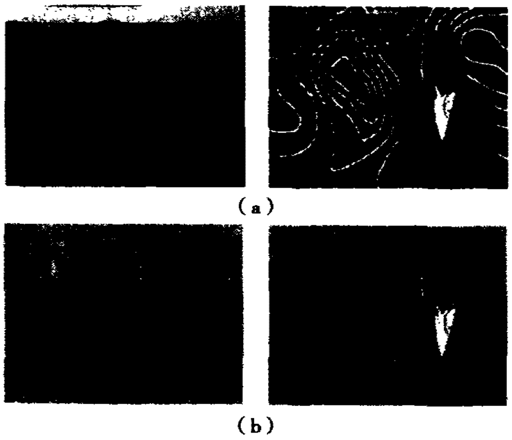

Fig. 2 The segmentation results of silent and weather forecast
(a) original frames
(b) final segmented objects
图2 Silent 和 weather forecast 的分割结果
(a)原始帧
(b)最后的分割结果

using weaker rules. The remaining URs are bound to their neighboring BCs which have the smallest spatial differences from them. The BCs whose FPPs are above 90% are labeled as moving objects. Otherwise, they are assigned to background. The segmentation mask is regularized with morphological tools. To enforce temporal continuity, the segmentation is aligned with that of the previous frame.

3 Experiment results and conclusions

The proposed segmentation approach is tested on several MPEG - 4 test sequences. Fig. 1 shows the four stages of the segmentation of Akiyo and Mother & Daughter sequences. The original frames are illustrated in Fig. 1(a) and Fig. 1(b) shown the segmented object regions by HOS test. Fig. 1(c), Fig. 1(d) and Fig. 1(e) are the detected object regions after stage 2, stage 3 and stage 4 are performed respectively. The segmentation results of two another sequences of Silent and Weather are demonstrated in Fig. 2, where Fig. 2(a) are the original frames and Fig. 2(b) are the final object regions. In all of the experiments, parameters of D, λ_1, λ_2 and λ_3 are set as 8, 10, 0.2 and 30 respectively.

It is shown in the experimental results that the proposed approach based on region binding performs well in motion objects segmentation. The region binding technique brings about flexibility and computational efficiency. Further investigation into automating the selection of thresholds is now under way.

REFERENCES

[1] Kim M, Choi J G, Lee M H, *et al.* Description of Core Experiments N2. *Doc. ISO/IEC JTC1/SC29/WG11*, San Jose, 1998

[2] Zucker S W. Region growing: Childhood and adolescence, Comput. Graph. *Image Process*, 1976, **5**(2): 382—399

[3] Horowitz S L, Pavlidis T. Picture segmentation by a directed split-and-merge procedure. *Proceeding of* 2^{nd} *Int. Joint Conference on Pattern Recognition*, 1974, 424—433

[4] Horowitz S L, Pavlidis T. Picture segmentation by a tree traversal algorithm. *J. Assoc. Comput. Mach.*, 1976, **23**(2): 368—388

[5] Chen S, Lin W, Chen C. Split-and-merge image segmentation based on localized feature analysis and statistical tests. *CVGIP: Graph Models Image Process*, 1991, **53**: 457—475

[6] Besl P, Jain R. Segmentation through variable-order surface fitting. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1988, **10**(1): 167—192

[7] Beveridge R. Segmenting images using localized histogram and region merging. *CVGIP'* 89, 1989, **2**(2): 311—347

[8] Hairs K. Hybrid image segmentation using watersheds and fast region merging, *IEEE Trans. Image Processing*, 1998, **7**(11): 1684—1699

[9] Adams R, Bishof L. Seeded region growing, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1994, **16**(4): 641—647

[10] Pavlids T, Liow Y T. Integration region growing and edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1990, **12**(1): 225—233

[11] Kwok S H, Constantinides A G. A fast recursive shortest spanning tree for image segmentation and edge detection. *IEEE Trans. Image Processing*, 1997, **6**(2): 328—332

[12] Morris O J, Lee M J, Constantinides A G. Graph theory for image analysis: An approach based on the shortest spanning tree. *Proc. Inst. Elect. Eng.*, 1986, **133**(1): 146—152

[13] Salembier P, Pardas M. Hierarchical morphological segmentation for image sequence coding. *IEEE Trans. Image Processing*, 1994, **3**(4): 639—651

[14] Neri A, Colonnese S, Russo G, *et al.* Automatic moving object and background separation. *Signal Processing*, 1998, **66**(2): 219—232

[15] Neri A, Colonnese S, Russo G. Automatic moving objects and background segmentation by means of Higher Order Statistics. *IS&T Electronic Imaging'* 97, *SPIE: Visual Communication and Image Processing*, S. Jose, 1997, 8—14