ISSN: 0976-9102(ONLINE)
DOI: 10.21917/ijivp.2014.0128

VIDEO SEGMENTATION USING A NOVEL LBP DESCRIPTOR

Zhongkun He[1], Yongping Zhang[2] and Dechun Zheng[3]
School of Electronic and Information Engineering, Ningbo University of Technology, China
E-mail: [1]hzk@nbut.cn, [2]ypz@nbut.cn, [3]nbzhengdc@126.com

Abstract

Video segmentation is the basis for content-based video retrieval, object recognition, object tracking, and video compression. This paper proposes a kind of novel and easy spatial-temporal LBP coding method, using the spatial-temporal 2 × 2 × 2 neighborhood clique to encode the changes in a video. Based on the coding method, a scheme of video segmentation is developed. Compared to the traditional segmentation method, its distinguished advantage is that it does not need to construct the background model and is simple in computation. Experimental results indicate that this new algorithm can give satisfying segmentation results.

Keywords:

Video Segmentation, Local Binary Patterns, Change Detections, Morphological Filters

1. INTRODUCTION

Video segmentation is very important in some application areas such as human-computer interaction, object-based video compression, and multi-object tracking. It is to segment the meaningful foreground object from the video.

Traditionally, there are some usual segmentation algorithms, which can be classified into three types: edge information based algorithm, image segmentation based algorithm and change detection based algorithm. Edge information based algorithm is to find edge information of each frame and keep tracking these edges. Then connect edge information to generate final object masks [1, 2, 3]. Though this algorithm is applicable under the situation of still camera and moving camera, its computation load is very large. Image segmentation based algorithms first apply image segmentation algorithms to separate a frame into many homogeneous regions. Then regions with motion vectors different from the global motion are merged as foreground regions [4, 5]. This algorithm can generate good segmentation, but the computation load is also high. Change detection based algorithms [6] threshold the frame difference to form change detection mask. Then the change detection masks are further processed to generate final object masks. Recently, Dynamic texture based algorithms have been proposed to analyze video [7, 8, 9, 10]. This kind of methods integrates spatial-temporal information to model the video.

This paper proposes a new LBP code to describe the change in the video. The proposed coding algorithm, which integrates the spatial-temporal information, is obviously simple. Because it can form the changes of video by coding, it is used to do the video segmentation. The results show it can effectively process the video segmentation.

This paper is organized as follows. In section 2.1 and 2.2, we describe the basic LBP code and spatial-temporal 2D LBP code. A new method of LBP code is presented in section 2.3. In section 3, we draw the scheme of this algorithm, choose the threshold and do post processing. The experimental results are shown in section 4. Finally, section 5 gives a conclusion of this paper.

2. LBP DESCRIPTOR

The local binary pattern (LBP) operator is first introduced by Ojala [14]. It is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. Instead of trying to explain texture formation on a pixel level, local patterns are formed. Each pixel is labeled with the code of the texture primitive that best matches the local neighborhood. In [9] and [13], the authors introduce the spatial-temporal LBP descriptor, which has XY, YT and YT planes to describe the video. These three planes use the 2D LBP descriptor to describe the changes of the video. We call it spatial-temporal 2D code. In this paper, we also use the LBP descriptor to analysis video, but directly in 3D encoding ways.

2.1 SPATIAL 2D LBP DESCRIPTOR

The concrete idea is that set some pixel as a center, compare its gray level with the gray level of those neighborhood pixels with certain radius and equal space interval, then get a group of binary number as its binary pattern. By calculating these numbers, a LBP code is generated by the following formula.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{p-1} s(g_p - g_c)2^p \tag{1}$$

and

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases} \tag{2}$$

where, (x_c, y_c) and g_c represent the spatial coordinates and gray level of the center pixel, respectively, g_p is the gray level of the p^{th} neighbor pixel, P and R correspond to number of neighbor pixels and coverage radius. Fig.1 shows an example for LBP code extraction with $P = 8$, $R = 1$.

5	4	3
4	3	1
2	0	3

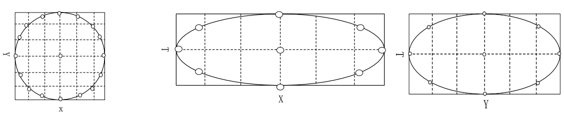
→

1	1	1
1		0
0	0	1

→

	143	

Fig.1. LBP code extraction



(a). XY Plane (b). XT Plane (c). YT Plane

Fig.2. Neighborhood cliques for coding in planes XY, XT, and YT

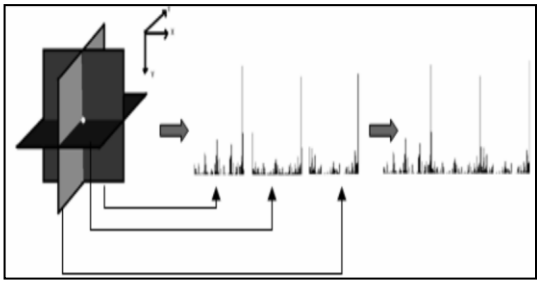


Fig.3. Spatial-temporal 2D LBP code

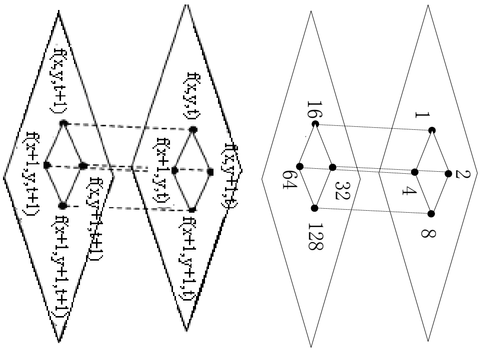


Fig.4. Neighborhood cliques and coding template

2.2 SPATIAL-TEMPORAL 2D LBP DESCRIPTOR

As explained in [13], the spatial-temporal 2D LBP code is generated by three planes which include XY, XT and YT. It incorporates spatial domain information and two spatial-temporal co-occurrence statistics together. Then, these three LBP codes, XY-LBP, XT-LBP and YT-LBP, are concatenated together. The neighborhood cliques and the coding process are shown in Fig.2 and Fig.3, respectively.

The essence of this approach is to use projection method describe the whole motion. The problem with the spatial-temporal 2D LBP comes from the time needed to compute the plane LBP codes at each step.

2.3 PROPOSED SPATIAL-TEMPORAL 3D LBP DESCRIPTOR

For detecting the motion in a video, the proposed spatial-temporal 3D LBP descriptor is based on a 2 × 2 × 2 neighborhood clique. Denote by $f(x, y, t)$ the pixel with coordinates (x, y) in the t^{th} frame. The neighborhood clique and template for coding are shown in Fig.4.

The concrete idea is: set some pixel as a center, compare its gray level with the gray level of three neighborhood pixels in this current frame and three corresponding pixels in the previous frame, then using Eq.(3) to extract the LBP code of the center pixel. In order to effectively detect changes, the sign function is modified as the form of formula Eq.(4). In this way, if there is not motion under the background with uniform distribution, the LBP code will be lower.

There is an important difference between the spatial-temporal 2D LBP and the spatial-temporal 3D LBP used in the proposed approach. As was explained in subsection 2.2, the spatial-temporal 2D LBP uses projections to describe overall motion. In the approach proposed here, the whole motion is characterized directly based on 3D neighborhood clique.

$$
\begin{aligned}
LBP(x,y,t) = {} & 2^1 \times s(f(x,y+1,t) - f(x,y,t)) \\
& + 2^2 \times s(f(x+1,y,t)) - f(x,y,t) \\
& + 2^3 \times s(f(x+1,y+1,t) - f(x,y,t)) \\
& + 2^4 \times s(f(x,y,t-1) - f(x,y,t)) \\
& + 2^5 \times s(f(x,y+1,t-1) - f(x,y,t)) \\
& + 2^6 \times s(f(x+1,y,t-1) - f(x,y,t)) \\
& + 2^7 \times s(f(x+1,y+1,t-1) - f(x,y,t))
\end{aligned}
\tag{3}
$$

$$
s(x) = \begin{cases} 1, & x > 0 \\ 0, & otherwise \end{cases} \tag{4}
$$

3. VIDEO SEGMENTATION ALGORITHM

After detecting the motion in a video, a segmentation scheme of moving targets can be developed by combining with thresholding techniques and morphological filtering. In this paper, the segmentation scheme consists of three main steps as shown in Fig.5: extracting LBP codes, binarizing, and morphological filtering.

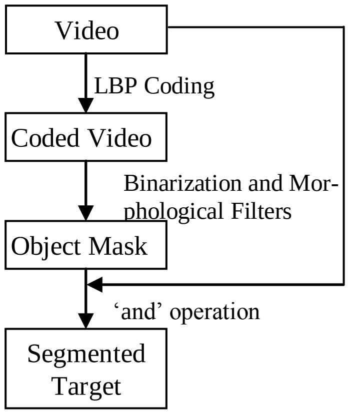


Fig.5. Video segmentation scheme

3.1 POST PROCESSING

Video with certain change information is got after using the novel LBP to process the original video. As shown in Fig.6(a), it gives the coding result of the gray-level frame image. Then choose a series of thresholds to binarize the frame image. In this paper, the iterative threshold mode is chosen to get rid of the noise as shown in Fig.6(b). The concrete idea is:

1) Choose the mid-value of the frame image as the initial threshold T;
2) Segment the frame image to 2 regions R_1 and R_2 using the threshold T, and calculate the mean value μ_1 and μ_2 in these two regions respectively taking use of Eq.(5).

$$
\mu_1 = \frac{\sum_{i=0}^{T} i n_i}{\sum_{i=0}^{T} n_i}, \quad \mu_2 = \frac{\sum_{i=T}^{L-1} i n_i}{\sum_{i=T}^{L-1} n_i} \tag{5}
$$

3) Use Eq.(6) to calculate a new threshold T_{i+1}:

$$T_{i+1}=\frac{1}{2}(\mu_1+\mu_2) \quad (6)$$

4) Repeat steps (2) and (3), until the difference between T_{i+1} and T_i is less than a given value. Set it as the final threshold.

The morphological closing operator is applied to the binary frames for producing the initial object masks. In order to remove the small holes in each initial mask, the seed propagation method is employed to generate connected object mask. Finally, the 'and' operator is applied to the original video and the generated masks, frame by frame to obtain the segmentation results.

The Fig.6 shows the process of creating the object mask using proposed method. The coding result for the 200^{th} frame of Claire is shown in Fig.6(a), the corresponding binarizing result is shown in Fig.6(b) and Fig.6(c) shows the created object mask.

3.2 EXPERIMENTAL RESULTS

The results of segmentation experiment with Claire video is shown in Fig.7. The top row shows the five frame images with gray level from the video. The object masks and segmented results are shown in the middle row and the bottom row, respectively.

For comparison, the frame difference algorithm is also used to segment the Claire video. As shown in Fig.8, it can be seen that the frame difference algorithm cannot get the shape of whole foreground object, but the proposed method can do that. Experimental results indicate that the proposed approach can achieve reasonably good segmentation.

(a) (b) (c)

Fig.6. Example of creating object mask: (a). LBP code of the 200^{th} frame of Claire, (b). Binarizing result, (c). Object mask

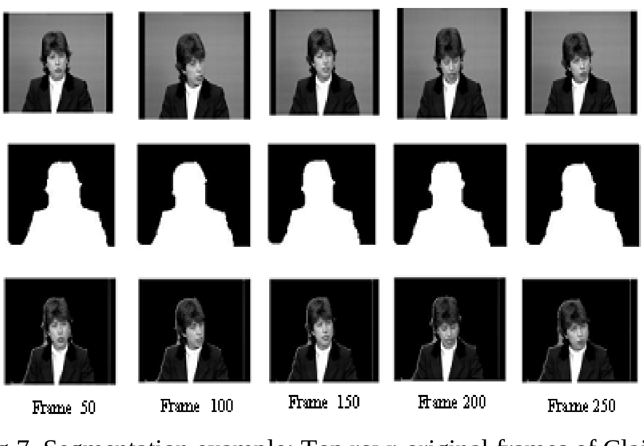


Fig.7. Segmentation example: Top row: original frames of Claire sequence, Middle row: object masks, Bottom row: segmented results

(a) (b) (c)

Fig.8. Comparative experiments: (a). 240^{th} original frame image, (b). Results of the 240^{th} based on frame difference segmentation, (c). Results of the 240^{th} based on the method proposed in this paper

4. CONCLUSION

In this paper, a spatial-temporal 3D LBP descriptor is proposed for detecting the motion in a video. The descriptor uses the spatial-temporal $2 \times 2 \times 2$ neighborhood clique, instead of three plane neighborhood cliques, to encode the video. Based on the proposed motion detection method, a video segmentation scheme is also developed. The experimental results demonstrate that the proposed algorithm can be effectively applied to video motion detection and moving object segmentation.

ACKNOWLEDGEMENT

This work is supported by the NSFC of Zhejiang Province, China (LY13F020014).

REFERENCES

[1] J. Canny, "A Computational Approach to Edge Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 6, pp. 679-698, 1986.
[2] C. Lü, D. Yuan and Q. Zhang, "Stereoscopic video object segmentation based on depth and edge Information", *International Symposium on Photoelectronic Detection and Imaging*, 2007.
[3] E. Sifakis, I. Grinias and G. Tziritas, "Video Segmentation Using Fast Marching and Region Growing Algorithms", *EURASIP Journal on Applied Signal Processing-Image Analysis for Multimedia Interactive Service-Part I*, Vol. 2002, No. 4, pp. 379-388, 2002.
[4] Q. Deng, H. Liu and L. Wu, "Video Object Segmentation Algorithm Based on Watershed and Region Merging", *Modern Scientist Instruments*, Vol. 2, No. 1, pp. 34-38, 2010.
[5] R. J. Radke, S. Andra, O. Al-Kofahi and B. Roysam, "Image Change Detection Algorithms: A Systematic Survey", *IEEE Transactions on Image Processing*, Vol. 14, No. 3, pp. 294-307, 2005.
[6] B. U. Toreyin, Y. Dedeoglu and A. E. Cetin, "Flame detection in video using hidden Markov models", *IEEE International Conference on Image Processing*, Vol. 2, pp. II-1230 -11-1233, 2005.
[7] G. Doretto, A. Chiuso, Y. N. Wu and S. Soatto, "Dynamic textures", *International Journal of Computer Vision*, Vol. 51, No. 2, pp. 91-109, 2003.
[8] A. B. Chan and N. Vasconcelos, "Layered Dynamic Textures", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 10, pp. 1862-1879, 2009.
[9] G. Zhao and M. Pietikäinen, "Local Binary Pattern Descriptors for Dynamic Texture Recognition", *18th Interna-*

tional Conference on Pattern Recognition, Vol. 2, pp. 211-214, 2006.

[10] J. N. Kapur, P. K. Sahoo and A. K. C. Wong, “A new method for gray-level picture thresholding using the entropy of Histogram”, *Computer Vision, Graphics and Image Processing*, Vol. 29, No. 3, pp. 273-285,1985.

[11] C. P. Yasira Beevi and S. Natarajan, “An efficient Video Segmentation Algorithm with Real time Adaptive Threshold Technique”, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 2, No. 4, pp. 13-28, 2009.

[12] N. Otsu, “A threshold selection method from gray-level histogram”, *IEEE Transactions on System, Man and Cybernetics*, Vol. 19, No. 1, pp. 62-66, 1979.

[13] G. Zhao and M. Pietikäinen, “Dynamic Texture Recognition Using Volume Local Binary Patterns”, *Dynamical Vision-Lecture Notes in Computer Science*, Vol. 4358, pp.165-177, 2007.

[14] T. Ojala, M. Pietikäinen and T. Mäenpää, “Multiresolution Gray Scale and Rotation Invariant Texture Analysis with Local Binary Patterns”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 971-987, 2002.

[15] P. L. Rosin, “Thresholding for change detection”, *Computer Vision and Image Understanding*, Vol. 86, No. 2, pp. 79-95, 2002.