A STEREO VIDEO OBJECT SEGMENTATION ALGORITHM BASED ON MOTION DETECTION AND DISPARITY

Lingyun Wang, Zhaohui Li, Dongmei Li

Department of Telecommunication System
Communication University of China, Beijing 100024, China
matey5210@cuc.edu.cn

Abstract: With the development of the stereo video, the technology of the stereo video object segmentation has gradually become a research hotpot. In this paper, a stereo video object segmentation algorithm based on motion detection and disparity is proposed, which utilizes the information of motion, disparity and edging features to extract moving objects from video sequences. Firstly, the eight-neighbor motion detection is used to get the change detection mask (CDM) of moving objects. Secondly, the stereo matching is adopted to get the disparity map which will then be used to modify the CDM. Finally, the edging information and mathematic morphology post-processing are applied to get the accurate video object. Experimental results demonstrate that the proposed algorithm can successfully extract moving objects from stereo video sequences whose background have other moving objects.

Keywords: Stereo video; Motion detection; Disparity estimation; Object segmentation

1 Introduction

Because of the development of multimedia technology, the information of multimedia are more wanted. Just using multimedia technology to browse can't meet consumers' demand, for they are calling for interactive features based on objects. As a result, the technology of video object segmentation came out.

Applications like video retrieval and video surveillance are evolved from 2D to 3D; therefore the stereo video object segmentation has become an important research topic. In the monocular video object segmentation algorithm, the accuracies of object segmentation results are highly subject to the limited information of monocular video sequences, but for stereo video object segmentation, the disparity information which are not available in the monocular video can be applied to improve the accuracies of object segmentation. The disparity information can better reserve the continuity inside the objects and the edging features of objects, so in the stereo video object segmentation, a relatively simple method can be employed to solve the problems that are difficult to solve in the monocular video object segmentation. Therefore the technology of the stereo video object segmentation is absolutely an active area of research.

For stereo video sequences, using single information such as color, edging, motion or depth can't obtain an ideal segmentation result. So ZhongJie Zhu and his partners in Ningbo University used contour tracking algorithm based on segmentation over monocular video sequences to extract the pair of objects in stereo video [1]. This algorithm can shorten the running time, but it is extremely sensitive to the accuracy of initial contour. JunGong Han and his partners in Shanghai University developed a disparity estimation algorithm based on delaunay triangulation with epipolar constraint [2], which is more accurate in disparity estimation but is slow in running time. ShiGang Wang and his colleagues in Jilin University put forward an algorithm combining the strong points of disparity map with frame difference, where firstly disparity map is segmented to get the object segmentation template in different disparity lays, and then frame difference extraction is made in the template regions to modify the object edge[3].

It's not hard to see that there are already many achievements in the study of stereo video segmentation, but there are still no common and efficient solutions. Take video sequences whose backgrounds have some other moving objects as an example, if the sequences are monocular, it will be very hard to extract the moving objects we need; but if the sequences are binocular, disparity or depth information which don't exist in monocular video can be used to segment the moving objects, because moving objects in the foreground and background have different disparity information. As a consequence, our paper proposed a stereo video object segmentation algorithm based on motion detection and disparity, which can successfully extract moving objects from stereo video sequences whose background have other moving objects.

2 Stereo video segmentation algorithm

2.1 Basic ideas of the algorithm

The basic ideas of the algorithm are listed as follows:

Firstly, our algorithm applies the eight-neighbor motion detection to get the change detection mask (CDM) of moving objects.

978-1-4673-2204-1/12/$31.00 ©2012 IEEE

Secondly, the algorithm adopts the stereo matching to get the disparity map which will then be used to modify the CDM.

Finally, in order to get a more accurate moving objects, edging information and mathematic morphology post-processing are used. The segmentation scheme is depicted in Figure 1.

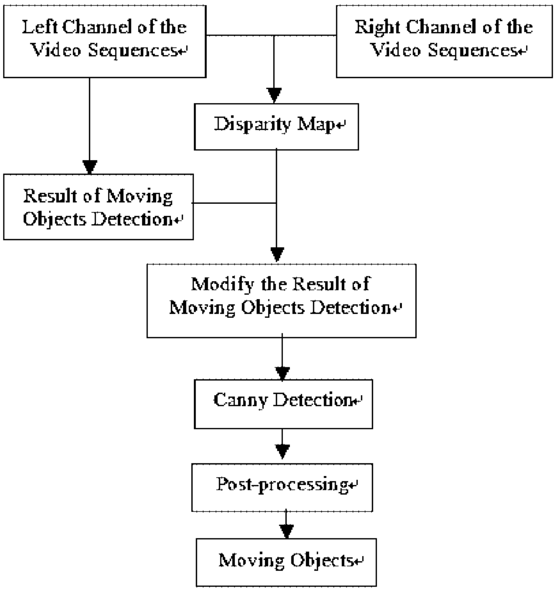


Figure 1 Flow chart of the algorithm

2.2 Motion detection

Motion detection exploits Change Detection Mask(CDM) instead of motion fields, which detects two adjacent frames to distinguish motion and static areas. Motion detection can be used to obtain the motion ranges of objects. This paper applies the eight-neighbor motion detection [4] to get the CDM of the moving objects' motion ranges. Firstly, the threshold estimated with noise characteristic parameters is used, and then the eight-neighbor motion detection is adopted to get the CDM.

Usually, the distribution of noise can be assumed a Gaussian distribution. Let $n_k(x,y)$ present the relative noise of the Kth and $(K+1)$th frame, then the $n_k(x,y)$also obeys Gaussian distribution.

$$n_k(x,y) \sim N(\mu, \sigma^2) \tag{1}$$

According to the "3 σ principle" in Gaussian distribution, the probability of a Gaussian variable lying in the range (X-3 σ .X+3 σ) is 99.73%, even though the range of Gaussian random variable is (-∞,+ ∞). Therefore, while $|d(x,y)-\mu|<3\sigma$, the $d(x,y)$ are defined as relative noise and are filtered. According to the "3 σ principle", to filter noise more precisely, we extend the 3 σ range to T, the noise decision principle is defined as:

$$\text{CDM}=\begin{cases}\text{Background} & |d(x,y)-\mu|<T \quad T\in[4\delta,6\delta]\\ \text{Moving Objects} & \text{Otherwise}\end{cases} \tag{2}$$

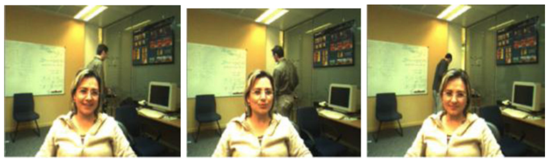
But the CDM extraction according to Eq. (2) is inaccurate and has many interior holes. In order to solve this problem, the eight-neighbor motion detection algorithm is applied, which is used to smooth the mask boundary and fill the interior hole. Let $d(x,y)$ represent the grey level difference mask, and the window M is selected centering in (x,y), then the maximum value T_{max} in the eight-neighbor regions M is defined as:

$$T_{\max} = \underset{(x,y\in M)}{MAX}(d(x,y)) \tag{3}$$

So Eq. (2) can be redefined as:

$$CDM=\begin{cases}\textit{Moving Objects} & |d(\mathrm{x,y})| \le T \wedge T\max > T\\ \textit{Moving Objects} & |d(\mathrm{x,y})>\mathrm{T}|\\ \textit{Background} & \textit{Others}\end{cases} \tag{4}$$

CDM results from Eq. (4) are the preliminary results of motion detection. Figure 2 shows the motion detection results of the left channel of the typical stereo video sequences *IU*. We can make out the motion ranges of moving objects, but there are also other moving objects in the background, so moving objects in the foreground can't be segmented effectively.



(a) Original frame 30th, 50th, 80th

(b) Preliminary segmentation results 30th, 50th, 80th

Figure 2 Preliminary segmentation results using moving detection

2.3 Disparity map estimation

Assume a calibrated stereo system, in this system matching points reside on corresponding horizontal lines. The disparity is calculated as the distance of these points when one of the two images is projected onto another. The disparity values for all the image points constitute the disparity map. Let F and B denote the focal length and baselines distance of two cameras, P_l and P_r denote the projections of 3D point P to the left and right image planes, (x_l, y_l) and (x_r, y_r) represent the coordinates of P_l and P_r, then the disparity vector can be defined as:

$$d = (x_l - x_r, y_l - y_r) \tag{5}$$

In the calibrated stereo system, $y_l = y_r$, so the relationship between disparity d and depth z can be simplified as:

$$d = x_1 - x_2 = BF\frac{1}{z} \quad (6)$$

According to (6), we know that if B and F are known, depth values can be worked out with disparity values.

The disparity map is usually achieved using the stereo matching algorithm. Stereo matching algorithm has two kinds of method: feature-based and region-based [5]. Feature-based method employs features such as points or lines to match feature points. These feature points are quite insensitive to noises, so matching is relatively accurate, but the disparity map obtained in feature-based method is sparse, which is not usable to modify the motion ranges of moving objects. But region-based matching method can generate a high density disparity map [6], which can meet the need of segmentation, though it is not quite accurate in occluded and low texturing areas,

In the calibrated stereo system, right images can be seen as left images offsetting along the horizontal axis, which means that if a point in the *Nth* line of the left image had a matching point in the right image the matching point must be in the *Nth* line of the right image. In addition, when stereo matching is carrying on from the left image to the right image, we are actually finding the matching points for the left images in the right images. So the disparity vector in horizontal orientation is negative, which implies that stereo matching can be limited in the horizontal and negative orientation, and matching range of values is between 0 and the max disparity value.

Our algorithm is based on the sum of absolute difference(SAD), with which the disparity of each pixel is generated. The definition of SAD is as followed:

$$SAD(x,y) = \sum_{m=-win}^{win} \sum_{n=-win}^{win} \left| f_L(x+m, y+n) - f_R(x+m-j, y+n) \right| \quad (7)$$

The value of *win* depends on the size of window, and j denotes disparity vector whose value is between 0 and the max disparity value. $f_L(x+m, y+n)$ and $f_R(x+m-j, y+j)$ represent the value of pixel in the left and the right images. The value of j which makes the SAD minimum is the disparity value of pixel (x,y) in the left image. The SAD algorithm doesn't need multiply operation, and is easy to operate.

The detailed processes in the disparity acquisition of pixel (x,y) in the left image are as followed:

1) Taking pixel (x,y) in the left image as the center point, create a M*M window, M=2*win+1.
2) Cover the same areas with the M*M window centered with pixel (x,y) in the right image.
3) Calculate the SAD value of pixel (x,y) according to formulation (7).
4) In the range between 0 and the max disparity value, move the window of the right image in the horizontal and negative orientation by pixel, which means the center of the window in the right image is (x-j, y),and then repeat operation (2) and (3).
5) Find the value of j which makes the SAD value minimum, and then find the matching point of pixel (x,y) in the left image. The matching point is pixel (x-j, y) in the right image, and the value of j denotes the disparity of pixel (x,y) in the left image.

Figure 3 shows the disparity map of sequences IU. Moving objects in the foreground and those in the background are very different in disparity, so the difference can be used to modify the preliminary results of motion detection. But at the same time, the generated disparity map exists many noises that may affect the modified results. Our algorithm applies a low-pass filter to get rid of the noise.

(a)Original frame 30th, 50th, 80th

(b) Disparity map 30th, 50th, 80th

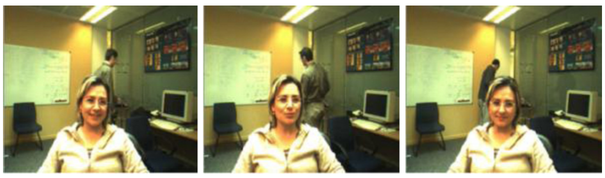
Figure 3 Disparity map results

2.4 Moving objects detection and post-processing

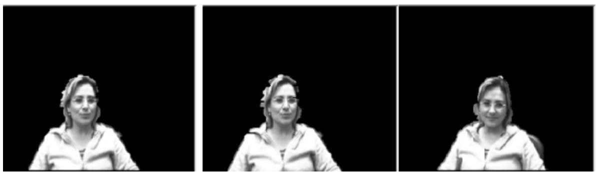
Moving objects we need to segment usually have small values in depth, according to the inverse relationship between depth and disparity, our algorithm sets threshold to get the probable range of moving objects. "And" operation is applied between the probable range generated by disparity map and the preliminary results of motion detection, which will modify the preliminary results. But the modified results still have some unclear and burring edging, interior holes. To solve this problem, this algorithm uses Canny detection to obtain the edging information, and then mathematic morphology post-processing is used to fill the interior holes in the moving objects. At last the method of dilation and erosion is applied to get the accurate video objects.

3 Experiment results

Using typical stereo video sequences *IU* to test the validity and efficiency of algorithm, the results are shown in Figure 4.



(a) Original frame 30th, 50th, 80th



(b) Final segmentation 30th, 50th, 80th

Figure 4 IU sequence segmentation

Figure 4 shows that proposed algorithm can successfully extract moving objects from stereo video sequences whose background have other moving objects.

4 Conclusions and discussions

In this paper, a stereo video object segmentation algorithm based on motion detection and disparity is proposed. The segmentation results show that the proposed algorithm can effectively solve the problem of extracting moving objects from stereo video sequences whose background have other moving objects. Our algorithm performs well in the situation that the distance among moving objects in the foreground and background is relatively long. In other situation, our algorithm is still acceptable if the accuracy in disparity is improved. therefore, in order to make the proposed algorithm be applied to more kinds of stereo video sequences, our future work is mainly focused on improving the accuracy of disparity map obtaining.

References

[1] ZhongGie Zhu,GangYi Jiang,Mei Yu,RangDing Wang,XunWei Wu. New Algorithm for Extracting and Tracking Moving Object in Object-Based Video Coding. Acta Electronica Sinica.2003-31(9).

[2] JunGong Han,Zhaoyang Lu,A Novel Disparity Estimation Algorithm for Stereo Image Encoding. Chinese journal of computers. 2003-26(12)

[3] Zhou qian，Wang Shigang，Yang Hong. Research on disparity-based object segmentation in stereo [C] / / Proceeding of the Asia-Pacific Workshop on Visual Information Processing. Beijing: 2006: 115 — 118.

[4] LingYun Wang,Zhaohui Li. An improved video object segmentation algorithm based on background rebuilding. The 2nd International Conference on Computer and Management.

[5] MASSICOTT J F，WILLSON S D，WYATT R，et a1. 1 480nm pumped erbium doped fiber amplifier with all optical automatic gain contml[J]. Electron Lett，1994，30(12)：962…964.

[6] An Ping，Zhang Zhao-yang，Ma Ran. Hierarchical MRF / GRF model based disparity estimation and segmentation for stereo images[J]. ACTA Electronic Sinica，2003，31(4)：597—601