

Research on Video Object Segmentation Method

ZHANG Wei

Baicheng Normal University
Baicheng, 137000, JiLin, China
zw19751016747@sina.com

Keywords: video segmentation; video object; motion estimation; object tracking

Abstract. With the advent of video standard MPEG-4 and MPEG-7, the video motion object segmentation becomes extremely important. The paper firstly introduced some basic concepts and the common methods of the video object segmentation technology, and then analyzed the advantages and disadvantages of each method, finally, forecasted the future and development trend of the technology.

Introduction

The human beings have access to the information society, and about eighty percent of the information perceived by human is visual information. Therefore, the video is related to the visual information which includes static image and time-varying image. The characteristics of static image are that information density change with space distribution instead of time; while the space density of time-varying image varies with time, therefore time-varying image is a time-space density mode, which can be represented by $f(x, y, t)$, x, y means spatial variable, t means time variable. The video of this paper refers to time-varying image.

Video segmentation refers to the division of the image or video sequence according to a certain standards for the purpose of defining some meaningful and relatively independent sets of space or time and then organizing the elements of these sets. It is not easy to make video object segmentation. The inherent problem about the formation of video segmentation is that the interesting video object is not about the low-level similarity of some items, such as color, brightness, and light flow etc.. In fact, the division of the video object contains the higher levels of semantic concept. So the methods of traditional video segmentation are not a meaningful division from the point of semantics. Therefore, how to use the right formula to express the semantic concept in segmentation is the main difficulty. So far, there is no algorithm can automatically divide the video sequences accurately and reliably.

Video object segmentation

Development of video segmentation technology: It is necessary to build video objects model before the video objects segmentation; however, there is not an accurate and generally accepted video object model nowadays, only a fuzzy definition which described as the meaningful entity that is linked together by each of the areas. But “meaningful” entity itself is a fuzzy concept. This is the reason why the video segmentation on any object is a very difficult research subject. Therefore, it is not easy to establish an exact model because of its different application in research; furthermore, the meanings of different video objects vary in different application. Video segmentation generated by objects is an important application domain in digital video processing technology, and the technology of digital video processing not only covers nearly every aspect of digital video processing and digital image processing and analysis but relates to computer vision, statistical signal processing, stochastic process and pattern recognition and other domains.

Method of video object segmentation: Video object segmentation is based on the static image segmentation. The most important point that differs from the static image segmentation lies in the introduction to motion information. The principle of video object segmentation mainly makes use of

motion information and separates the foreground object from its background by combining with information about other colors, textures, edges, regions and so on. The following briefly introduces some common methods and also analyzes some advantages and disadvantages of them.

Video segmentation technology based on the temporal information: Temporal segmentation can identify the motion object and its method mainly includes change detection, optical flow method, object tracking, motion estimation, etc.. Change detection is a common method of detecting moving object. Its basic principle is that the current video frame is segmented into “changed” and “unchanged” area relative to reference frame through the detection of frame difference between the former and latter frame, unchanged part represents static background, changed part represents motion and occlusion. Using change detection for video object segmentation is simple in principle, but there are the following problems: First, threshold selection for calculating the detection template is very important, but at present there is no general method. Second, change detection for some texture is not full and the video sequence for some frame objects in the presence of motion while others in the absence of motion is invalid. Third, the identification between the occlusion area of object and the visible area is also a difficult problem.

Another one is Optical flow method. When human observes movement objects, object image forms a series of consecutive change images on the retina, the series of change information is continuously “flowing” the retina, like a light “flow”, so it is called optical flow. Optical flow expresses image changes, including target motion information, thus can be used to determine the target’s motion. The segmentation based on optical flow actually obtains approximate athletic field from a sequence of images through the study of the optical flow field, and then makes video segmentation according to characteristics of the motion field. The optical flow method is a common method used to estimate the motion field; the optical flow method applies projection velocity model distinguishing from projection displacement model. But there is the occlusion/exposed problem for the method and also steady assumption of optical flow field can not be satisfied at the point of the brightness mutation, which limits the application of optical flow method.

Motion estimation is an estimation process of the reference frame of the pixels in the current frame. Motion estimation techniques rely on two assumptions: one is that lighting is constant on the object movement track, and the illumination light does not change with time when the object moves so as to ensure the image brightness pattern change to be caused by the movement rather than light alterations. The other is no shading background. Although these assumptions are not sufficient to gain real world video sequences, most motion estimation methods are based on those assumptions. Motion vector field usually has two kinds of descriptive methods: non-parametric representation and parameter representation. Nonparametric algorithm estimates dense motion field so that each pixel is assigned a corresponding flow vector and the aperture problem is solved by increasing the smoothness constraint conditions to make neighbor pixels have similar flow vector. But the introduction of smoothness constraints caused the movement of edge blur, motion segmentation algorithm using the estimated motion field alone will not get the accurate segmentation results, therefore, nonparametric motion field is generally not used directly for segmentation except the flat shift. Parameters model is to describe the 3D surface movement (displacement and velocity) in the image plane of an orthographic or perspective projection. Parametric motion description first needs to make the scene segmentation. Each region in the pixel motion can use a set of parameters to represent; each motion vector of the pixel can be integrated by a set of parameters. Parametric motion description is not sensitive to noise, as long as it is consistent with regional boundaries, there will be no moving boundary ambiguity. The disadvantage of parameter expression is the need for segmentation and often attaching strict assumptions to the scene and motion.

Video segmentation technology based on Spatio-temporal joint: Spatial information and temporal information separately corresponds to the video intra-frame and inter-frame information. In terms of a complete sequence, they are only partial information and reflect some of characteristics of the video scene. Therefore, there are some limitations on them. Spatial segmentation can accurately reflect the object boundary information, but due to the lack of movement information, it can’t individually segment to extract moving objects segmentation from the scene. Only using the motion

information to segment the target region is often not very accurate and segmentation region boundary is not accurate, furthermore, the target area is not continuous, empty exists in the inner of the target area, and isolated noise area also exists. Therefore, we must accurately segment video objects and make full use of spatial information and temporal information. Segmentation method of Spatio-temporal joint is based on spatial and temporal clustering segmentation method, based on the tracking or judgment method of region theory, based on object tracking method and so on.

Segmentation method based on the clustering theory of space-time pixel. Among which, each pixel is usually a vector to describe its characteristics, all of the vector contains the information such as the pixel location, color, motion vector and so on. The similarity between the different pixels is measured by the differences between the vectors, then the spatial clustering of the pixel is measured based on this similarity. Pixel clustering method did not take full advantage of the boundary distinctions in the space domain information and the correlations within the region, so the partition of the boundary is often not precise enough. In addition, among the pixel clustering methods, the process of space-time joint analysis and the process of the space-time joint segmentation are not separate. This makes this kind of methods only applicable to certain time-domain detection means.

Segmentation method based on region. Regional binding method is based on such an assumption: Each target area is spliced by gray scale consistent with the region. This assumption makes segmentation boundary of regional binding method more accurate in many cases. According to the before and after order relation between the spatial segmentation and temporal segmentation, regional binding method is divided into two classes, that is, regional clustering method and proportion method of the uniform area. Regional clustering method is to merge and cluster different gray regions based on a certain similarity metric. First, divide the current frame into some consistent grayscale areas and then calculate the motion model parameters for each region and you can use a threshold to define whether the movement model parameters of the two regions are the same or not in accordance with the motion model's consistency with the adjacent consistent gray scale region. The combined area will re-calculate the motion model parameters; and repeat the above merging process until there are no any areas that can be combined. Under the condition of having known the number of moving targets, the optical flow field is smoothly constraint by using the same color region, and then the k-means clustering is made for the smoothed optical flow field to determine the different movement target areas. Regional clustering method has two key issues: First, it requires that the segmentation of consistent grayscale areas is entirely correct, not across the two different target areas at the same time, but the fact is often not so good; Second, since the motion vectors and limited estimation accuracy of regional movement models is sometimes difficult to have similar motion model parameters to determine whether the two adjacent consensus regions belong to the same moving target based on two consistent grayscale regions. Proportion method is the short form for regional proportion method. the principle is: if the region R^i has pixels of more than a certain

percentage ρ marked as j ($j \in \Gamma$) in F , the pixels of the entire R^i region are labeled as j , that is, thinking the entire R^i belong to the j target. Among which, R^i is a grayscale consistent region. All R^i sets represent segmentation results of consistent grayscale region, $\Gamma = \{0, 1, 2 \dots L-1\}$ as a symbol set, L is the number of the target area, if we only segmented motion regions from the background, $\Gamma = 1$ (0 for background, 1 for the target). The proportion method has advantages of simple operation, good flexibility. In addition, under the condition of more accurate segmentation of consistent grayscale region, proportion method inherited the advantages of accurate segmentation boundary from region binding method. These advantages make it most commonly used joint space-time analysis method. However, the proportion method over-relied on the accuracy of the segmentation of consistent grayscale region and choices of thresholds ρ . Object tracking can be understood as the process of searching for the video object in subsequent frames. Object tracking is commonly used model matching. Common tracking methods include active contour model, Hausdorff object tracking and so on.

Video segmentation techniques based on the Hierarchical Structure: By minimizing the quadratic determine method, the method fitted the parameter model to the entire change area from one frame to the next frame, and then divided this area into continuous small areas based on goodness fit between a single model and each region or sub-regions, and the flow chart of the algorithm is shown in Fig. 1:

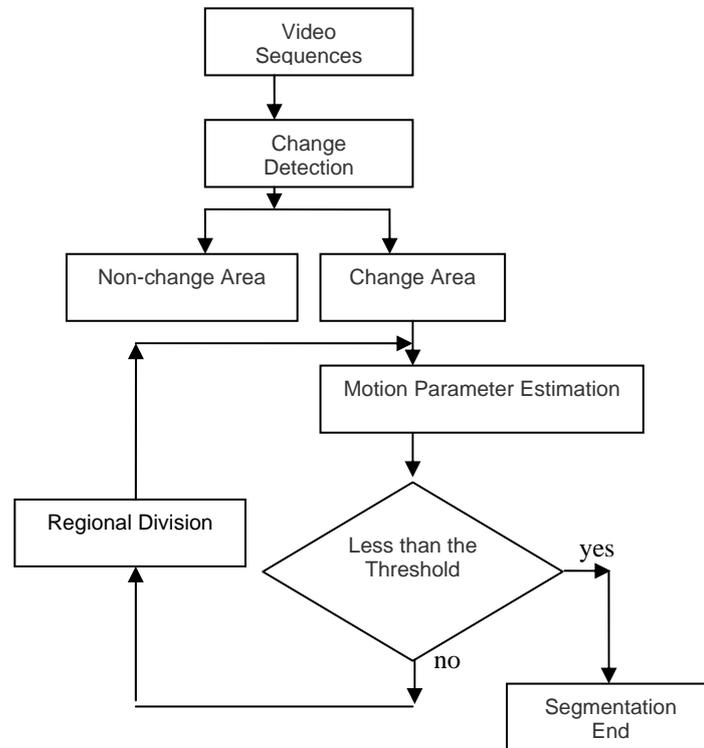


Fig.1 Video Segmentation Based on the Hierarchical Structure

Step1, use change detection to detect changed region between the adjacent two frames, and consider the change region of the connectivity of each space as a moving object.

Step2, estimate the motion parameter models for each moving object, such as the affine parameter model, the eight-parameter model and so on, thus use the model parameters to show the motion characteristics.

Step3, under the parameter model of each estimated motion region, calculate the average value of the displaced frame difference of pixels within each motion region, if the average value of the displaced frame difference is zero or less than a predetermined threshold, the estimated motion parameter model is considered to be effective, otherwise it is invalid. Mark invalid areas of each motion parameter, and then further divided these areas, repeats the processes of step1 and step2 until motion parameter models for each area are effective.

This kind of method can be seen as a top-down hierarchical method and the motion parameter model varies in different literatures. The algorithm does not consider the boundary information of the object, if the boundary information is given in the above algorithm, it will further enhance the accuracy of moving object's boundary for the final segmentation.

Evaluation criterion of video segmentation algorithm

Although the human eye can easily distinguish the video object, but the technology of extracting video object depending on the computer is still in the immature stage of development. Since evaluation criterion of video object segmentation algorithm is no uniform standard, generally speaking, the following aspects will mainly be considered:

Segmentation quality: Video motion object segmentation can make use of minimal interaction, automatic, accurate segmentation of motion object boundary.

Computational complexity: It has the advantage of a small amount of calculation, and video moving object segmentation can be applied in general machine, fast by the software without reliance on high-end computer.

Algorithm generality: Video object segmentation is made without a prior knowledge for the video object color, shape, motion, type.

Artificial interaction level: It needn't too much manual interaction.

Flexibility: Allows flexibility in video motion object segmentation is allowed to interactively correct some errors under the condition of the error occurs.

Conclusion

The biggest difficulty of video object segmentation comes from two aspects: on the one hand, it is difficult to establish a uniform prior model due to the extreme complexity and diversity of scenes in the real world, which makes it difficult to find a single method to complete all of the video object segmentation tasks; On the other hand, for high level semantic video object, it is very difficult to extract through the low level visual segmentation at present, so image analysis techniques by using machine language to accurately define and describe should be further developed. Video object segmentation is a difficult point in computer vision and video processing, but this technology actually has good prospect of application. In addition, the technology itself with highly challenge is accompanied by the infinite charm, attracting countless dedicated explorer to make continuous research.

References

- [1]CHU Guang-li, "Video Object Segmentation Algorithm Based on Background Reconstruction", International Conference on Computer Design and Applications, Hebei, 2010
- [2]ZHANG Wei, "Brief analysis of Formatting Standard of Isomerism Data Converting", International Conference on Electronics, Communications and control, 2011
- [3]Lijun Ma, "The Integrity Rules and Constraints of Database" Proceedings of 2011 International conference on Electronic and Mechanical Engineering and Information Technology vol.7:3747-3749
- [4]Guowen CHE, Lijun MA, Liyuan XU, "An Excel Model of Analogue Orthotropic Plate Method to Calculate Load Transverse Distribution Coefficients", Proceedings of 2011 International conference on Electronic and Mechanical Engineering and Information Technology vol.7:3796-3799
- [5]WANG Jun ,SHA Yun,WU Yu-shu, "Automatic Video Segmentation Method Based on Background Model", Computer Engineering and Applications 2003
- [6]BAO Hong-qiang, " Content-based Video Moving Object Segmentation Technology", [Doctoral Dissertation] Shanghai: Shanghai University 2005
- [7]YUAN Ji-wei,SHI Zhong-ke, "The Background Extraction Algorithm of a Fast Moving Target",Application Research of Computers 2003.