Video Image Segmentation and Object Detection using Markov Random Field Model and EM Algorithm

Prof. Abhinav V. Deshpande
Assistant Professor
Department of Electronics and telecommunication Engineering
Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati-444701

Abstract - With many application in various domains, object detection and tracking has received a great deal of attention over the decades in the field of image analysis and computer vision. It has been studied by scientists from different areas of psychophysical sciences and those from different areas of computer science. However, developing a computer algorithm to do the same thing is one of the toughest tasks in computer vision. Research over the past several years enables similar such as foreground detection, detecting connecting regions, extracting object features. Correspondence-based object matching, detecting left and removed objects. This research paper gives a review of different object detection techniques which are available as of today. The focus is on the segmentation techniques by investigating the use of image frames. The motion based, spatiotemporal, temporal are under segmentation techniques. We incorporate the EM algorithm with these segmentation techniques for better performance of detecting and tracking of an object in this research paper.

Key words: Object Detection and Tracking, Segmentation Techniques, MRF Model, EM Algorithm

I. INTRODUCTION

Detection and Tracking of moving objects from a video scene is a challenging task in video processing and computer vision [1][2][3][4]. It has wide applications such as video surveillance, event detection, activity recognition, activity based human recognition, fault diagnosis, anomaly detection, robotics, autonomous navigation, dynamic scene analysis, path detection and others etc. [1][2][3][4]. Moving object detection in a video is the process of identifying different object regions which are moving with respect to the background. More specifically, moving object detection in a video is the process of identifying those objects in the video whose movements will create a dynamic variation in the scene [2]. This can be achieved by two different ways:

1) Motion Detection/Change Detection
2) Motion Estimation

Change or motion detection is the process of identifying changed and unchanged regions from the extracted video image frames when the camera is fixed and the objects are moving. For motion estimation, we compute the motion vectors to estimate the positions of the moving objects from frame to frame. In case of motion estimation, both the objects and the camera may move [2]. After detecting the moving objects from the image frames, it is required to track them. Tracking of a moving object from a video sequence helps in finding the velocity, acceleration and position of it at different instants of time. In visual surveillance, sometimes it may be required to obtain the speed/velocity of a moving vehicle so as to keep an eye on the movement of a particular vehicle [6]. Moving object detection by the process of motion/change detection is again restricted by the requirement of a reference frame (where the object is not present). This can be accomplished by the use of intensity difference based motion detection algorithm [4] (where objects may move slow or fast). In the absence of a reference frame, if there is a substantial amount of movement of an object from one frame to another the object can be tracked exactly by generating a reference frame [4]. However for those cases where the reference frame is not available and

3) the objects in the scene do not have a substantial amount of movement from frame to frame or
4) the objects in a given scene move and stop for some time and move further, identification of moving objects becomes difficult with temporal segmentation [1][2][3][4]. A robust video image segmentation algorithm is essential to solve these problems. Watershed algorithm (a region based approach) [1][2][3][4] is a famous approach in this context. A computationally efficient watershed based spatial segmentation approach was proposed by Salember et. al. [5]. They had used spatial segmentation and temporal segmentation to detect the object boundaries. However, this method produced over segmented results and hence could not detect the objects satisfactorily.

II. RELATED WORK

Several tracking methods have been proposed in the literature. The main non-mutually exclusive categories which are identified are given as below: Region-based tracking, active contour-based tracking, feature-based tracking, model-based tracking and body part-based tracking. The method which is used [7] proposes a tracking algorithm that predicts the object by predicting the object boundary by using block motion vectors which are followed by updating the contour by using occlusion/ disocclusion detection. The method in [8] uses background subtraction and motion estimation for tracking an object. DCT domain background subtraction in the Y plane is used to locate the candidate objects in the subsequent I-frames after a user has marked an object of interest in the given frame.

By using Cb and Cr planes and motion vectors are used to select the target object from the set of candidate objects. Another group of algorithms deal with object tracking by using adaptive particle filters [9], Kalman Filter [10]. The literature [11] provides a critical review of the compressed domain indexing techniques. The method in [9] proposes a tracking algorithm which is based on pixel features in the wavelet domain. Two stage object tracking [12] is performed by combining the region-based method and the contour-based method. The method in [13][14] novel algorithm for moving object detection and tracking. The proposed algorithm includes two schemes: one for spatio-temporal spatial segmentation and the other for temporal segmentation.

III. SEGMENTATION TECHNIQUES

JSEG is one of the segmentation techniques for identifying and tracking an object. The essential idea of JSEG is to separate the segmentation process into two independently processed stages, color quantization and spatial segmentation. In the first stage, colors in the image are quantized to several representing classes that can be used to differentiate the regions in the image. By this way, it is easy to identify an object. A region growing method is then used to segment the image based on the multi-scale J images. The goal is to achieve consistent segmentation and tracking results, even for the scenes with arbitrary non-rigid object motion [6]. In the JSEG technique, the image or an object is segmented by using the j-values, so with this we cannot get equal and small segments of an object to avoid this we go for spatial and temporal segmentation methods.

IV. J-SEGMENTATION

It has been attempted to address the moving object detection by using the method of temporal segmentation. It was found that the temporal segmentation could help to construct the video object plane (VOP) and detect the objects. In all these cases, it was assumed to have the reference frames. This scheme produced poor results when the video has slow moving objects and Diehl's algorithm. The algorithm is implemented in four steps which are given as below:

[1] An initial change detection mask (CDMi) between two successive frames is generated by thresholding the difference image. The boundaries of the CDMi are smoothed by a relaxation on a MAP detector by using local thresholds which consider the state of neighboring pixels. This results in a change detection mask (CDM). The CDM is simplified by using a morphological closing operator and elimination of small regions. An initial moving object mask (Omi) is estimated by eliminating the uncovered background from the CDM. At the last step, the CMi is adapted to the luminance edges of the corresponding frame, resulting in the final object mask (OM).

V. SPATIAL SEGMENTATION

The spatial segmentation splits the entire image into homogeneous regions in terms of intensity. The different homogeneous regions are distinguished by their encompassing boundaries. The spatial segmentation algorithm is implemented in four steps. Choi et. al. segmentation scheme [CLK 97, m2091]

The input images (or motion compensated image if there are global motion) are simplified by the morphological open and close operations by using reconstruction filters. These filters remove the regions that are smaller than a given size but preserve the contours of the remaining objects in the image. The spatial gradient of the simplified image is approximated by the use of morphological gradient operator. In order to increase the robustness, the color information is also incorporated into the gradient computation and the estimated gradient is thresholded to remove the noisy gradients. The spatial gradient is used as an input of watershed algorithm to partition an image into homogeneous intensity regions. The boundary decision is taken through the use of watershed algorithm. The watershed algorithm is a region growing algorithm and it assigns the pixels in the uncertainty area to the most similar region with some segmentation criteria such as difference of intensity values. The watershed algorithm is highly sensitive to the gradient noise, which yields many catchment basins, the final result of the algorithm is usually an over segmented tessellation of the input image. In order to overcome this problem, the region merging followed. Finally we will merge all the segments in an order.

VI. SPATIAL TEMPORAL SEGMENTATION

Owing to the emerging multimedia applications such as MPEG-4 and MPEG-7 there is a need of segmentation of scenes into meaningful objects or meaningful moving objects in order to facilitate the so called content-based functionalities video object plane (VOP) in order to support the content-based functionalities such as object-based spatial and temporal scalability.

The JSEG method can be modified to segment he color and texture regions in the video data. Here, our approach differs from the existing techniques in that it does not estimate the exact object motion. The most common approaches involve the estimation of the motion vectors, whether based on optical flow or optical flow or affine motion matching [14]. The following are some of the common problems that are associated with motion estimation. The dense field motion vectors are not very reliable for noisy data. The affine transform is often inadequate when modeling the motion in close-up shots, especially when the objects are turning away from the camera. The occluded and uncovered areas introduce the errors in estimation.

VII. THE EXPECTATION-MAXIMIZATION (EM) ALGORITHM

The EM algorithm is a general technique for finding the maximum likelihood (ML) criterion which estimates with the incomplete data. In the classic ML [15], the estimate is obtained from the complete data as follows:

$$\Theta^* = \arg\max \ln P(dcom \mid \theta) \quad \ldots\ldots\ldots\ldots\ldots (1.1)$$

In EM, the complete data is considered to consist of the two parts dcom = {dobs, dmis} of which only dobs is observed while dm is missing (or unobservable and hidden). With only the incomplete data dobs an EM procedure attempts to solve the following ML estimation problem

$$\Theta^* = \arg\max \ln P(dobs \mid \theta) \quad \ldots\ldots\ldots\ldots\ldots (1.2)$$

which is more general than the classic ML.

Intuitively, we may envisage a maximum likelihood procedure that after some initialization for dmis and θ iterates between the following two steps until convergence.

1) Estimate the missing part as dm is given the current θ estimate and then it is used to augment the observed data d = dobs to from the complete data set dcom = {dobs, dmis} and
2) Estimate the θ, with dcom by maximizing the complete data log likelihood in P (dmis, dobs, θ). The simultaneous labeling estimation algorithm which is described in the previous two subsections is based on such an idea. However, we cannot work with this log-

likelihood directly because it is a random function of the missing variables f (the reason why the procedure above is adhoc). The idea of the EM algorithm is to use the expectation of the complete data log-likelihood, E ln P (dmis, dobs, θ). Related to the parameter estimation in MRF models, the missing part corresponds to the unobservable labeling f, f = dmis and the observed part is the given data, d = dobs. The complete data log-likelihood is then denoted by ln P (f, d | θ).

The EM algorithm consists of the following two steps for each iteration:

1) The E-Step:
   Compute the following conditional expectation of the log-likelihood
   $Q(\theta \mid \theta(t)) = E[\ln P(f, d \mid \theta) \mid d, \theta(t)]$……... (1.3)
2) The M-step:
   Maximize $Q(\theta \mid \theta(t))$ in order to obtain the next estimate

If in the M-step, the next estimate θ (t+1) is chosen only to ensure

The E-step computes the conditional expectation of the unobservable labels f as given in the observed data d and the current estimate θ (t) and then substitutes the expectations for the labels. The M-step performs maximum likelihood estimation as if there were no missing data (i. e. as if they had been filled in by the expectations)

This describes just one instance of the EM procedure. More generally, the probability P (f, d | θ) can be replaced by any complete data sufficient statistics. When the prior distribution p (θ) is known, the M-step can be modified in order to maximize the expected posterior

$E \ln p(\theta, f \mid d) \mid d, \theta(t)$ ………………..... (1.4)

where

$P(\theta, f \mid d) = p(f, d \mid \theta)\, p(\theta) / P(d)$………. (1.5)

instead of the likelihood.

VIII. CONCLUSION

The proposed scheme and the spatio-temporal spatial segmentation result of the initial frame is obtained by using edge-based MRF modeling and a hybrid MAP estimation algorithm (hybrid of SA and ICM). The segmentation result of the initial frame together with some change information from the other frames is used to generate an initialization for segmentation of other frames. Then, an ICM algorithm is used on that frame starting from the obtained initialization for segmentation. It is found that the proposed approach produces better segmentation results as compared to those of edgeless and JSEG segmentation schemes. The proposed scheme uses EM algorithm for better accuracy as compared to the considered MRF based segmentation schemes for a number of video sequences.

ACKNOWLEDGMENTS

I am very much thankful to all of the staff members and Head of Department, Department of Electronics & Telecommunication Engineering, Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati-444701 for their kind support and co-operation in successful carrying out this research work. This research work was undertaken as a part of improvement and expansion of the current and emerging trends in the field of Electronics & Telecommunication Engineering under the Technical Education Quality Improvement Program (TEQIP-2) so that the new and young generation of research scholars working in the field of research and development in the Digital Image Processing platform could get the benefit to expand their horizons and learn the new techniques of image segmentation through object detection algorithms like the one which we have used that is the Markov Random Field Model and the EM Algorithm in order to extract a specific region of interest from video images which are captured from a moving camera thereby exploiting the individual pixels of the test images and characterizing the distance between the two pixels of an image by means of EM algorithm.

REFERENCES

[1] A. L. Bovic, “Image and Video Processing: New York, Academic Press, 2000.
[2] A. M. Tekalp, “Digital Video Processing”, Englewood Cliffs, NJ:Prentice Hall, 1995.
[3] Y. Amit, “2-D Object Detection and Recognition”, Cambridge, MA:MIT Press, 2002.
[4] R. F. Gonzalez and R. E. Woods, “Digital Image Processing”, Singapore:Pearson Education, 2001.
[5] P. Salember and F. Marques, “Region-Based Representation of Image and Video Segmentation Tools for Multimedia Services”, IEEE Transactions on Circuits Systems Video Technology, Volume 9, No. 8, pp. 1147-1169, December 1999.
[6] B. Li and R. Chellappa, “A Generic Approach to Simultaneous Tracking and Verification in Video”, IEEE Transactions on Image Processing, Volume 11, No. 5, pp. 530-544, May 2002.
[7] Karthik Hariharakrishnan and Dan Schonfeld, “Fast Object Tracking Using Adaptive Block Matching”, IEEE Transactions on Multimedia, Volume 7, No. 5, pp. 853-859, October 2005.
[8] Ashwini Aggarwal, Susmit Biswas, Sanjeep Singh, Shamik Sural and A. K. Majumdar, “Object Tracking using Background Subtraction and Motion Estimation in MPEG Videos”, Springer-Verlag Berlinheidelberg, ACCV 2006, LNCS 3852, pp. 121-130.
[9] Xi Tao, Zhang Shengxiu and Yan Shiyuan, “A Robust Visual Tracking Approach with Adaptive Particle Filtering”, IEEE Second International Conference on Communication Software and Networks, 2010, pp. 549-553.
[10] Li Ying-Hong, Pang Yi-Gui, Li Zheng-Xi and Liu Ya-Li, “An Intelligent Tracking Technology Based on Kalman and Mean Shift Algorithm”, IEEE Second International Conference on Computer Modeling and Simulation, 2010.
[11] Madhur Mehta, Chandni Goyal, M. C. Srivastava and R. C. Jain, “Real Time Object Detection and Tracking: Histogram Matching and Kalman Filter Approach”, IEEE 2010, 978-1-4244-5586-7.
[12] M. Khansari, H. R. Rabiee, M. Asadi and M. Ghanbari, “Object Tracking in Crowded Video Scenes Based on the Undecimated Wavelet Features and Texture”, EURASIP Journal on Advances in Signal Processing, Volume 2008, pp. 692-698.

[13] Qiang Chen, Quan Sen Sun, Pheng Ann Heng and De-Shen Xia, "Two Stage Object Tracking Method Based on Kernel and Active Contour", IEEE Transactions on Circuits and Systems for Video Technology, Volume 20, No. 4, April 2010.
[14] S. D. Babacan and T. N. Pappas, "Spatiotemporal Algorithm for Background Subtraction", in Proceedings of IEEE ICASSP, April 2007, pp. 1065-1068.
[15] http://www.springer.com/series/4205

