International Journal of Research Studies in Science, Engineering and Technology
Volume 1, Issue 9, December 2014, PP 39-48
ISSN 2349-4751 (Print) & ISSN 2349-476X (Online)

Video Based Crowd Density Estimation and Prediction System for a Wide Area Surveillance

[1]Mamatha, [2]Ravikiran

[1]PG Student, Chalapathi Institute of Technology, Mothadaka, Guntur
[2]Assoc Professor, Chalapathi Institute of Technology, Mothadaka, Guntur

Abstract: *Visual surveillance in dynamic scenes, especially for human and some objects is one of the most active research areas. An attempt has been made to this issue in this work. It has wide spectrum of promising application including human identification to detect the suspicious behavior, crowd flux statistics, and congestion analysis using multiple cameras. In this paper deals with the problem of detecting and tracking multiple moving people in a static background. Detection of foreground object is done by background subtraction. Detected objects are identified and analyzed through different blobs. Then tracking is performed by matching corresponding features of blob. An algorithm has been developed in this perspective using Angular Deviation of Center of Gravity (ADCG), which gives a satisfying result for segmentation of human object.*

Keywords: *Tracking, Visual Surveillance, Blob, Center of Gravity (CG) and Feature Extraction.*

1. INTRODUCTION

As an active research topic in Computer Vision, Visual Surveillance in dynamic scenes attempt to detect, recognize and track certain object from image sequences and more generally to understand the human or any object behavior. The aim of this research is to develop an intelligent surveillance system for tracking human in dynamic scenes. It has wide range of potential applications such as security issue in important installation, traffic Surveillance in expressways, to measure the crowd flux in railway station, airports etc. In surveillance system considerable amount of work has been carried out by researchers [1].

Technology has reached a stage where video camera may be affordable in public and private areas [2] for keeping track of movement of any human or object. This paper has presented a vision-based system for accurate segmentation and tracking of moving objects in cluttered and dynamic outdoor environments, surveyed by a single fixed camera. Each foreground Object of Interest (OI) has been segmented and shadows/highlights removed.

The video surveillance system usually has two major components, one is detecting moving object the other one is to tracking them in sequence from video images. The accuracy of these components largely affects the accuracy of overall surveillance system. Detecting moving regions in the scene and separating them from background image is a challenging problem. In the real world, some of the challenges associated with foreground object segmentation are illumination changes, shadows, camouflage in color, dynamic background and foreground aperture [3]. Foreground object segmentation can be done by three basic approaches: frame differencing, background subtraction and optical flow. Frame differencing technique does not require any knowledge about background and is very adaptive to dynamic environments [4], but may suffers from the problem of foreground aperture due to homogeneous color of moving object. Background subtraction can extract all moving pixels, but it requires perfect modeling. It is extremely sensitive to scene changes due to lighting and movement of background object. Optical flow, one of the robust technique to detect all moving objects, even in the presence of camera motion, but it may be computationally expensive and may have limited application. Object can be represented as, *Points:* The object is represented by a point, In general, the point representation is suitable for tracking objects that occupy small regions in an image. *Primitive geometric shapes:* Object shape may be represented by a rectangle, ellipse, shown in Though the primitive geometric shapes are more suitable for representing simple rigid objects. However they may also be used to represent nonrigid objects.

2. PROPOSED SYSTEM

2.1. Video Tracking

The video surveillance system usually has two major components, one is detecting moving object the other one is to tracking them in sequence from video images. The accuracy of these components largely affects the accuracy of overall surveillance system. Detecting moving regions in the scene and separating them from background image is a challenging problem.

In the real world, some of the challenges associated with foreground object segmentation are illumination changes, shadows, and camouflage in color, dynamic background and foreground aperture. Foreground object segmentation can be done by three basic approaches: frame differencing, background subtraction and optical flow.

Frame differencing technique does not require any knowledge about background and is very adaptive to dynamic environments , but may suffers from the problem of foreground aperture due to homogeneous color of moving object. Background subtraction can extract all moving pixels, but it requires perfect modeling.

It is extremely sensitive to scene changes due to lighting and movement of background object. Optical flow, one of the robust technique to detect all moving objects, even in the presence of camera motion, but it may be computationally expensive and may have limited application. Object can be represented as, Points.

The object is represented by a point that is the centroid shown in (Figure 1(a)) In general, the point representation is suitable for tracking objects that occupy small regions in an image. Primitive geometric shapes: Object shape may be represented by a rectangle, ellipse, shown in (Figure 1(c), (d), etc. Though the primitive geometric shapes are more suitable for representing simple rigid objects. However they may also be used to represent no rigid objects.

Object silhouette and contour: Contour representation may be used to define the boundary of an object. The region inside the contour is called the silhouette of the object. Silhouette and contour representations are suitable for tracking complex nonrigid shapes or objects. Articulated shape models: Articulated objects are composed of body parts with different joints. For example, the human body is an articulated object with torso, legs, hands, head, and feet connected by joints. In order to represent an articulated object, one can model that constituent part by integrating different graphical shape like cylinders or ellipses.

2.1.1. Skeletal Models

Object skeleton can be extracted by applying medial axis transform to the object silhouette. This model is commonly used as a shape representation for recognizing objects. Skeleton representation can be used to model both articulated and rigid objects.

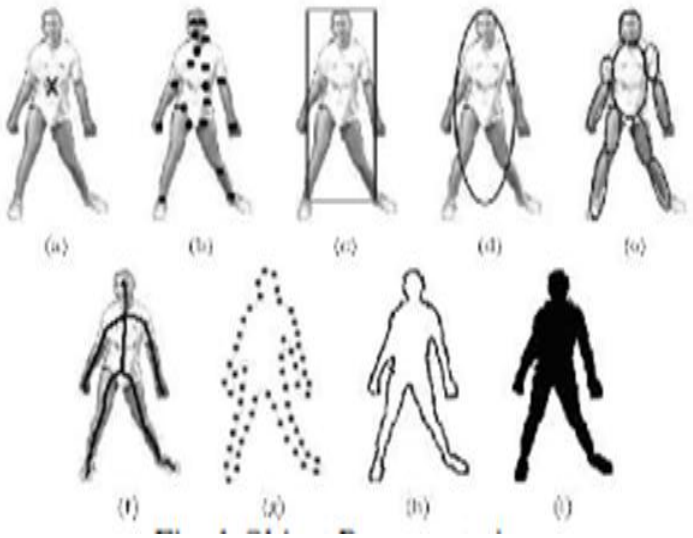

Fig6.1. *object Representation*

It is difficult to get a background model from the video because background information keeps always changing by different factors like illumination, shadows etc. The static background is considered for analyzing the object in this paper. Background subtraction method is used for detecting moving object, because it gives maximum number of moving pixels in a frame.

Object tracking methods usually divided into four groups [9], they are:

- Region-based tracking
- Active-contour-based tracking
- Feature-based tracking
- Model-based tracking

It is not so easy because of some of the problems, which generally occur during tracking. Occlusion handling problem i.e. overlapping of moving blobs has to be dealt carefully. Other problems like lighting condition, shaking camera, shadow detection, similarity of people in shape, color and size also pose a great challenge to efficient tracking.

It is based on binocular stereovision using prediction verification paradigm. Adaptive change in motion detection is performed to detect moving object in the scene. There are many reviews on image segmentation: Pal and Pal [5], which does not go details into the algorithms, but which classifies segmentation technique, discuss advantages and disadvantages of each class of the segmentation method and contain exhaustive list references to the literature up to the early 1990's.

2.2. Tracking

A feature-based object-tracking algorithm requires useful feature selection, feature extraction, feature matching and proper handling of object's appearance and disappearance. Object Entry and Exit in a scene was proposed by Stauffer. Most of the works on tracking use a prediction on features in the next frame and compare the predicted value with estimated value to update the model. Usually a model like Kalman filter is used for prediction. Techniques like Euclidean distance function successfully used by Xu, Collins et al. used a correlation function for matching regions in motion.

Comaniciu proposed a mean-shift technique to calculate most probable target position. They calculated similarity of objects by constructing histograms of target model and target candidates. Similarity is expressed by a metric derived from the Bhattacharyya coefficient.

2.3. Tracking System

Our surveillance activity goes through three phases. In first phase the target is detected in each video frame. Segmentation using background subtraction is generally used to identify any moving object in the scene, but some time due to some environmental factors such as light condition, camera position detected object are splitted into more than one blob. While acquiring target proposed methodology namely Angular Deviation of Center of Gravity (ADCG) which could be useful to combining the splitted blob and grouped into a single object.

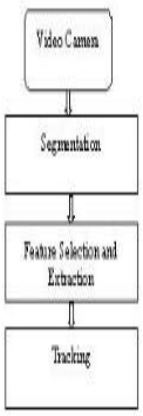

Fig. 2: Block Diagram of the Tracking System

2.4. Definition of Tracking

Automatic surveillance systems generally track moving objects from one frame to another in an image sequence. The tracking process goal is to associate to the moving objects found by the segmentation process information like its identity, position, speed and acceleration.

The ideal behaviors of a tracking system is to provide a set of tracks (possibly in an environment reference system) that are in a one-to-one correspondence with the objects appearing in the field of view of the sensor, i.e., no tracks associated with different objects and no multiple tracks associated with the same object. Factors that make this problem difficult include occlusion, complex light sources, and large size changes in the field of view .Two major components can be distinguished in a typical visual tracker:

- Targetrepresentationandlocalizationismostlyabottom-upprocesswhich has also to cope with the changes in the appearance of the target;
- Filtering and data association is mostly a top-down process dealing with the dynamics of the tracked object, learning of scene priors, and evaluation of different hypotheses.

The way the two components are combined and weighted is a the joint observational model relies on the representation of the targets. For example, if a people tracker is considered, the representation can assume the human body to be made of three parts: head, torso, and legs . The body is assumed to be a whole volumetric entity, described by its position in the 3D plane and having a given volume and appearance captured by color intensity values.

The joint observational model works by evaluating a separate appearance score for each object, encoded by a distance between the histograms of the model and the hypothesis (i.e., a sample), involving also a joint reasoning captured by an occlusion map.

The occlusion map is a 2D projection of the 3D scene which focuses on the particular object under analysis, giving insight on what are the expected visible portions of that object. This is obtained by exploiting the hybrid particles set {x }NK in an incremental visit procedure on the image plane p p=1.

- The first hypothesis to be evaluated is the nearest to the camera
- Its presence determines an occluding cone in the scene, where the confidence of the occlusion depends on the observational likelihood achieved;
- Particles in the scene that are farther from the camera and that fall in the cone of occlusion of other particles are less considered in their observational likelihood computation.

When a new observation is received, all existing tracks are projected forward to the time of the new measurement (predict step of the filter) and the observation is assigned to the nearest of such predicted state.

The distance between observations and predicted filter states is computed considering also the relative uncertainties (covariances) associated with them. The most widely used measure of the correlation between two mean and covari- ance pair {x1, P1} and {x2, P2}, which are assumed to be Gaussian-distributed random variables, is:

$$\exp(-1(x1 - x2)(P1 + P2) - 1(x1 - x2)T)\ 2$$

$$\text{Pass}(x1, x2) = \quad 2\pi \mid (P1 + P2) \mid$$

If this quantity is above a given threshold, the two estimates are considered to be feasibly correlated. An observation is assigned to the track with which it has the highest association ranking. In this way, a multiple-target problem can be decomposed into a set of single-target problems. In addition, when an observation is ”close enough” to more than one track, multiple hypotheses are generated (see Track split operation).

2.5. Track Management

Track formation. When a new observation is obtained, if it is not highly cor- related with any existing track, then a new track is created and a new Kalman filter is initialized with the position (x,y) observed and given to all the not observed components (e.g., velocity) a null value with a relatively high covari- ance. If the subsequent observations confirm the track existence, the filter will converge to the real state.

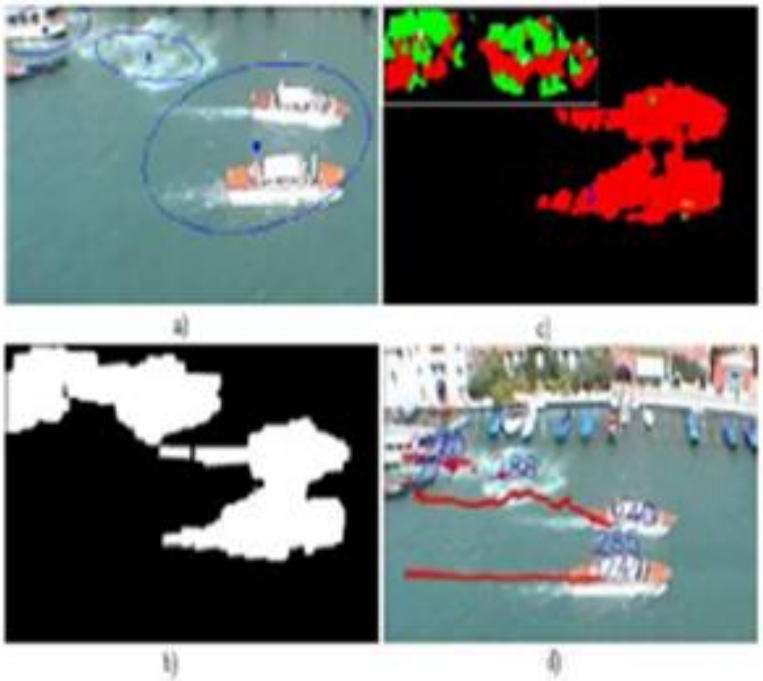

Fig7.2. *a) Two boats very near are detected as one because there is a b) Foreground under segmentation error and c) Optical flow does not solve the problem but d) With the multi-hypothesis method the system continues to track the boats separated over time.*

2.5.1. Track Update

Once observations are associated to tracks, standard update of the Kalman Filters are performed and the filters normally evolve.

Track split. When an observation is highly correlated with more than one track, new association hypotheses are created. The new observation received is used to update all the tracks with which it has a probability association that exceeds the threshold value. A copy of each not updated track is also maintained (track split). Subsequent observations can be used to determine which assignment is correct. This splitting is limited to the best 2 associations. Moreover, we limit the total number of hypotheses to 50.

2.5.2. Track Merges

This step aims at detecting redundant tracks, i.e. tracks that (typically after being split) lock onto the same object. At each step, for each track the correlation with all the other tracks is calculated using equation (9.1). If the probability association between two tracks exceeds a threshold (experi- mentally established), one of the two tracks is deleted, keeping only the most significant hypothesis.

Track deletion finally, when a track is not supported by observations, the uncertainty in the state estimate increases and when this is over a threshold, we can delete the track from the system. We have considered, as a measure of the uncertainty in the state estimate of each target, the Kalman filter gain relative to the track.

An example of the multi-hypothesis tracking method when two boats are very near they are detected as one because there is a foreground under-segmentation error and also the optical flow does not solve the problem because the two boats proceed in the same direction Thanks to the multi-hypothesis approach the system considers the wrong observation as a new track but it continues to track over time the former two because of the history of the observations.

After some frame, when the boats separate each other, the two correct tracks survive since they are supported by observations, while the new erroneous hypothesis will be deleted. Obviously, the multi-hypothesis tracker does not remove all the error cases, but it is successful in many cases in which a single-hypothesis tracker would fail.

3. SIMULATION RESULTS

3.1. Browsing

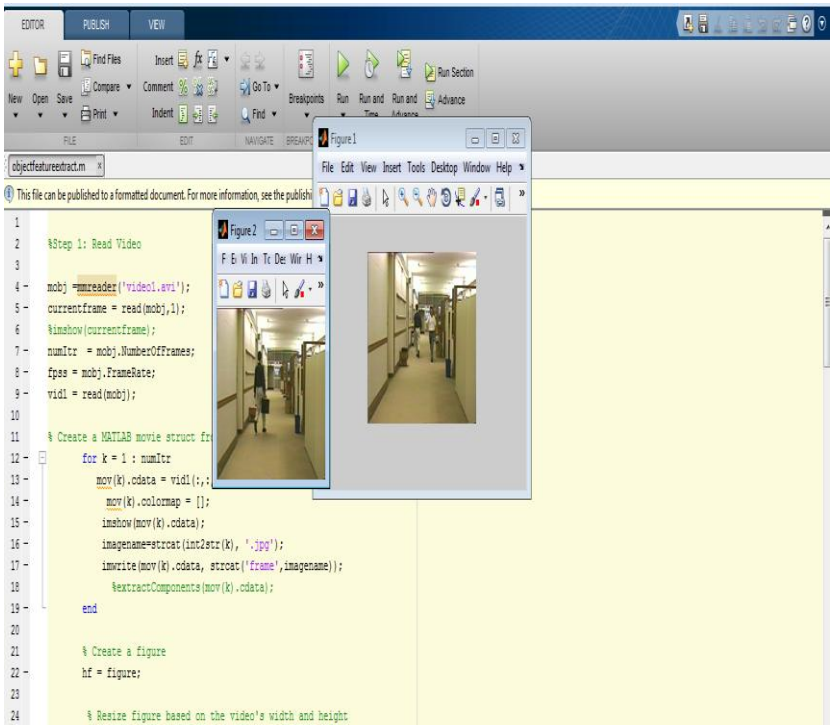

Fig10.1. *Video Browsing*

The above fig10.1 is video browsing input video. To read a video in the script we use different commands such as ‘mmreader’, ‘videofreader’.

3.2. Segmentation

Fig.9.2.Segmentation Output

The above fig10.2 is output of video segmentation. We can observe tiny movements of the persons.

3.3. Tracking Output

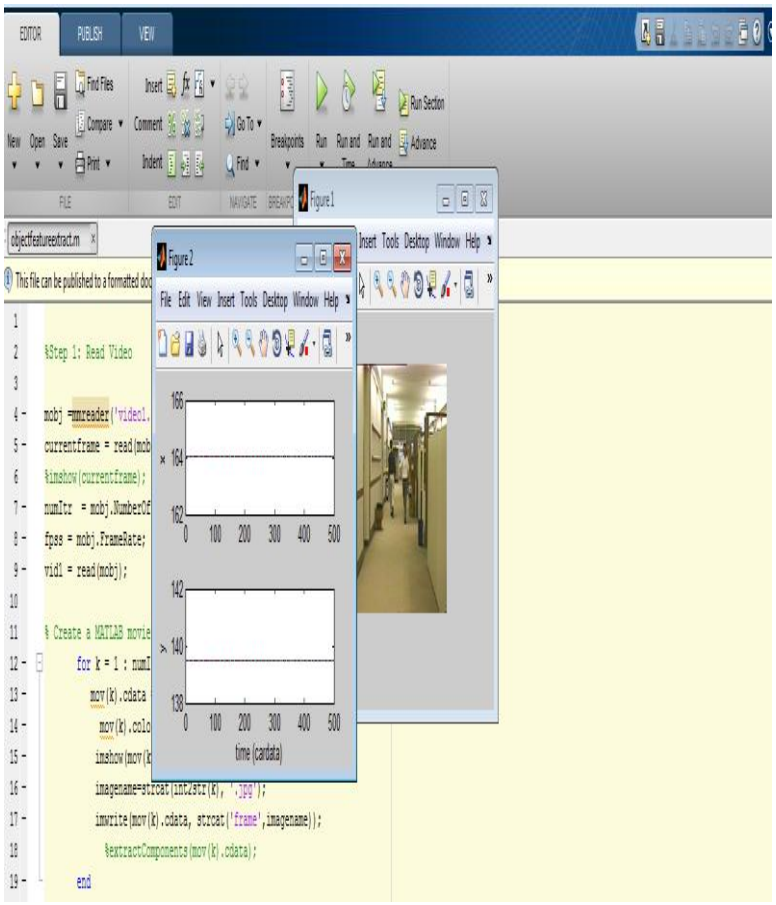

Fig10.3. *First stage of tracking*

The above fig10.3 is first stage of tracking indicates the moving objects and a graph is drawn between movement of people and time. The movement of people on x-axis and time is taken on y-axis.

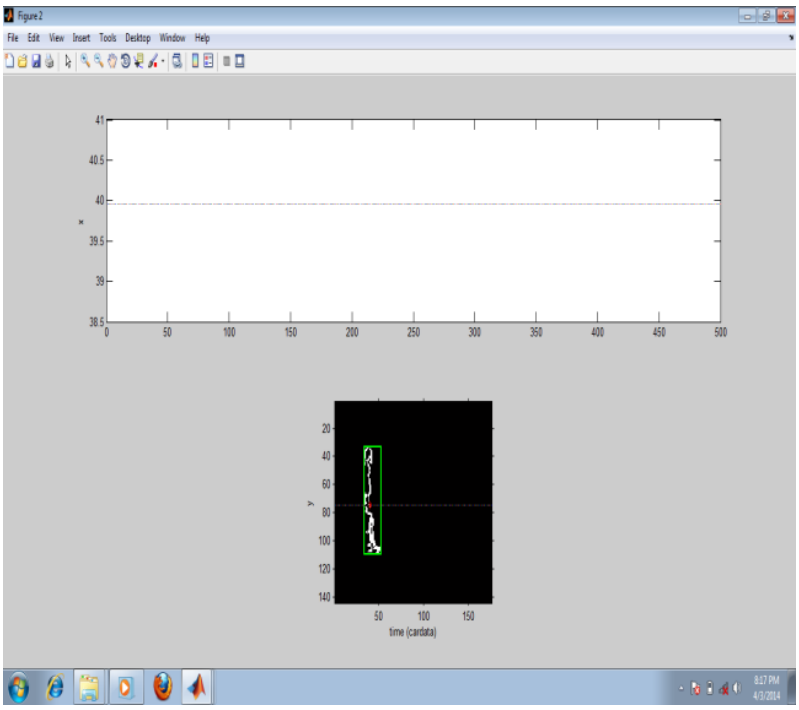

Fig10.4.*Tracking output 2*

The above fig10.1 is second stage of tracking in which the green color rectangular box indicates the movement of the person and we see his movements on the graph also.

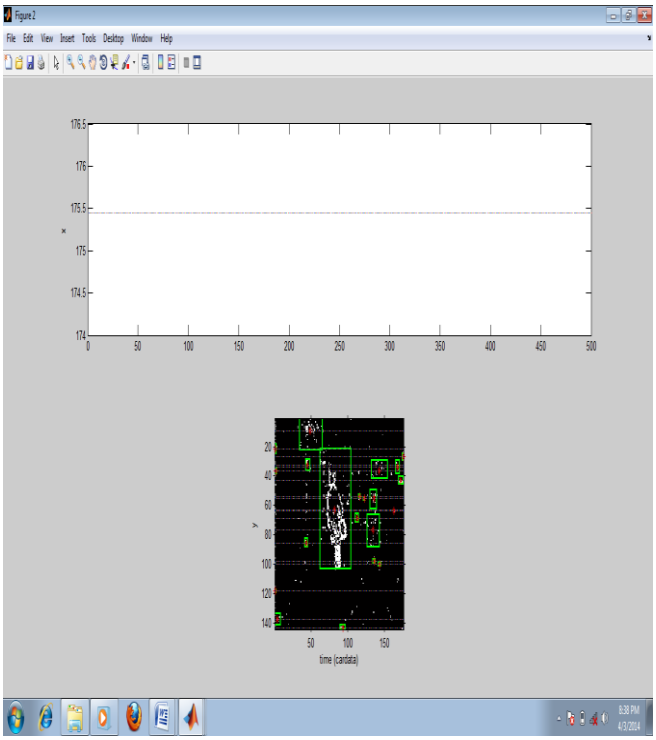

Fig10.4. *Tracking output 3*

The above fig10.1 is final stage of tracking in which all the rectangular boxes indicates the moving objects and their necessary movements and the graph is flucuvates.The graph changes with respective the movement of the crowd or person.

4. CONCLUSION

So far we have discussed about a crowd density estimation and prediction system for wide-area security. AMID based approach is applied to detect crowded areas and a geometry module is included to correct perspective distortion. The liner fitting method estimates the number of people in a crowd and the velocity is also obtained by the optical flow method. After crowd density and velocity are estimated, the prediction module is used to estimate the crowd density at designated points at a later time.

Compared to existing methods, the proposed method is a real time system for applications and the crowd density analysis algorithm can work properly in both low and high crowd density scenes. Experiments and real applications demonstrate the effectiveness and robustness of our method in real scenes although there are some aspects to be improved in the system. A block-based dense optical flow with spatial and temporal filtering is used to obtain velocities that can be used to infer the location of objects among crowded scenarios.

In this thesis automatic video surveillance systems have been analyzed, giving an overview of the state-of-the-art as well as proposing novel approaches and solutions .This chapter presents a review and comparative study of various topics in the area of crowd video analysis. The advantages and disadvantages of the state-of-the-art methods related to video analytics in crowded scenes have been detailed. Tracking individuals in a high-density crowd has been addressed in recent years, as opposed to previously tracking individuals in sparse or even ad-hoc scenarios.

A major advance is the introduction of high-level crowd motion pattern as a prior into a general framework [4, 63]. However, the problem of tracking still remains as a challenging problem in the area of computer vision. One major challenge for tracking in a crowded scene is inter-object occlusion due to the interactions of participants in a crowd. There remains a gap between the state-of-the-art and robust tracking of people in a crowded scene. Most recent trackers for crowds use Particle Filters, using different kinds of features; the use of self-similarity measures for this particular application can be of interest and deserves further research, given the results it achieved in other Computer Vision fields. During recent years there has been substantial progress towards understanding crowd behavior and abnormality detection based on modeling crowd motion pat- tern. However, these approaches

capture general movement of a crowd but do not accurately detect details of individual movements. As a result, the current literature in understanding crowd motion is not ready to capture the motion pattern of a un- structured crowd scene where the motion of the crowd appears to be random [63].

Future research in this area requires localized modeling of crowd motion to capture different behaviors in unstructured crowded scenes. On the other hand, the under-standing and modeling of crowd behavior remains immature despite the considerable advances in human activity analysis. Progress in this area requires further advances in modeling or representation of a crowd event and recognition of these events in a natural environment.

5. FUTURE SCOPE

Significant progress has been achieved over the last decade in the field of automatic video surveillance, but successful cases are limited to "controlled" situations, in which is possible to insert into the system strong knowledge about the environment to monitor (e.g., systems for highway monitoring).

The main difficulty in realizing effective video surveillance systems that can reliably work in real conditions is the need of implementing techniques that are robust to the many different conditions that arise in real environments.

Research challenges come with practical considerations such as the physical placement of cameras, the network bandwidth required to support them, installation costs, privacy concerns, and aesthetic constraints.

Today, automatic video surveillance systems can detect pre-programmed events such as abandoned luggage, intrusion, and overcrowding. They can pro- vide useful information about traffic statistics and people counting, but the challenge is in proving understanding at a higher level.As a short-term objective, the automatic video surveillance

REFERENCES

[1] ZHAN Beibei, MONEKOSSO D N, REMAGNINO P, et al. Crowd Analysis: A Survey [J]. Machine Vision and Applications, 2008, 19(5-6): 345-357.

[2] XU Liqun, ANJULAN A. Crowd Behaviors Analysis in Dynamic Visual Scenes of Complex Environment[C]// Proceedings of the 15th IEEE International Conference on Image Processing, 2008 (ICIP 2008): October 12-15, 2008. San Diego, CA, USA, 2008: 9-12.

[3] GUO Jinnian, WU Xinyu, CAO Tian, et al. Crowd Density Estimation Via Markov Random Field (MRF)[C]// Proceedings of 2010 8th World Congress on Intelligent Control and Automation (WCICA): July 7-9, 2010. Jinan, China, 2010: 258-263.

[4] MA Ruihua, LI Liyuan, HUANG Weimin, et al.On Pixel Count Based Crowd Density Estimation for Visual Surveillance[C]// Proceedings of 2004 IEEE Conference on Cybernetics and Intelligent Systems: December 1-3, 2004. Singapore, 2004: 170-173.

[5] PARAGIOS N,RAMESH V. A MRF Based Approach for Real-time Subway Monitoring[c]// Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.December814,2001.Kauai,HI,USA,2001:1034-1040.

[6] LIN S F,CHEN J Y,CHAO H X.Estimation of number of people in crowded scenes using perspective transformation.IEEE Transaction on system , man and cybernetics part

[7] SUBBURAMAN V B,DESCAMPS A, CARINCOTTEC.Couting people in the crowd Using a Generic Head Detector [C]// proceedings of 2012 IEEE 9TH International conference Advanced Video and signal-Based Surveillance (AVSS): September 18-21,2012Beijing,China,2012:470-75

[8] WU Xinyu ,LIANG Guoyuan,LEE K K,et al.Crowd Density Estimation using Texture Analysis and Learning [C]// Proceeding of IEEE international conference on Robotics and Biomimetric,2006(ROBIO'06):December 17 -20,2006.Kunming ,china,2006:214-219

[9] T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," IEEE Trans. on Syst., Man and Cybern., Part C: Applications anReviews, vol. 34, no. 3, pp. 334-352, 2004.

[10] R.T. Collins, A.J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson, "A System for Video Surveillance and Monitoring," The Robotics Inst. Carnegie Mellon Univ., CMU-RI-TR-00-12, 2000.

[11] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system forlearning statistical motion patterns," IEEE Trans. Pattern Anal. and Mach. Intell., vol. 28, no. 9, pp. 1450–1464, 2006.

[12] X. Wang, X. Ma, and E. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models," IEEE Trans. Pattern Anal. and Mach. Intell., vol. 31, no. 1, pp. 539–555, 2009.

[13] X. Wang, K. Tieu, and E. Grimson, "Trajectory analysis and semantic regions modeling using nonparametric hierachical bayesian models," Int'l J. Computer Vis, vol. 28, no. 9, pp. 1450–1464, 2011.

[14] I. Saleemi, L. Hartung, and M. Shah, "Scene understanding by statistical modeling of motion patterns," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010.

[15] E. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371-1375, 1998.

[16] S.S. Fels and G.E. Hinton, "Glove-talk: a neural network interface which maps gestures to parallel format speech synthesizer controls," IEEE Trans. Neural Network, vol. 9, no.1, pp. 205-212, 1997.