International Journal For Technological Research In Engineering
Volume 1, Issue 12, August-2014 ISSN (Online): 2347 - 4718

VIDEO OBJECT TRACKING BASED ON AUTOMATIC BACKGROUND SEGMENTATION: A SURVEY

Ravikiran G Banarase-Deshmukh[1], Parmalik Kumar[2], Prasad Bhosle[3]
[1,2,3]M. Tech Scholar, Asst. Prof, ME Scholar, Dept of CSE
[1,2]Patel Institute of Technology, Bhopal, India
[3]Dr.V.B.Kolte College of Engg, Malkapur, India

Abstract: Object tracking and segmentation is the ultimate purpose of many video processing systems. Higher level analysis and understanding of events require certain low level computer vision tasks to be performed. In this paper only on these low-level features are focus, whose achievement in turn determine the success of high-level investigation everyday jobs, like understanding the interactions between persons, recognizing and interpreting human body interaction, detect nonstandard actions and so on. The two critical, low-level computer vision tasks that have been undertaken in this work are: Foreground-Background Segmentation and Object Tracking. We expand the basic philosophy of Background Subtraction used widely for foreground-background segmentation. This involves the subtraction of the current image from a reference or estimated background image. The error image is then threshold to detect the foreground pixels. We use a stochastic model of the background and also adapt the representation through time. This adaptive nature is essential for long-term surveillance applications, mainly when the background composition or intensity distribution changes with time. In such cases, concept of a static reference background would no longer make sense.
Keywords: Object tracking, segmentation, video processing, and Background subtraction.

I. INTRODUCTION

Visual tracking is one of the most significant fields in video processing and computer vision has been widely applied to traffic surveillance system, suspicious person monitoring system etc. In practical purpose, since the camera moves and rotates, it needs to track objects in a dynamical background. How to select the initial target objects automatically and establish objects motion model, and how to update object and background models at each frame are the key in real-time Visual tracking with an active camera. Aimed at solving the problem of initial selection of object, Reference used optic flow field and internal limitation constrains of motion to detect motion object. Video surveillance systems have long been in use to monitor security sensitive areas. The history of video surveillance consists of three generations of systems which are called 1GSS, 2GSS and 3GSS. The first generation surveillance systems (1GSS, 1960-1980) were based on analog sub systems for image acquisition, transmission and processing. They extended human eye in spatial sense by transmitting the outputs of a number of cameras monitoring a set of sites to the displays in a central control room. They had the major drawbacks like requiring high bandwidth, difficult archiving and retrieval of events due to large number of video tape requirements and difficult online event detection which only depended on human operators with limited attention span. The next generation surveillance systems (2GSS, 1980-2000) were hybrids in the sense that they used both analog and digital sub systems to resolve some drawbacks of its predecessors. They made use of the early advances in digital video processing methods that provide assistance to the human operators by filtering out spurious events. Most of the work during 2GSS is focused on real-time event detection. Third generation surveillance systems (3GSS, 2000-) provide end-to-end digital systems. Image acquisition and processing at the sensor level, communication through mobile and fixed heterogeneous broadband networks and image storage at the central servers benefit from low cost digital infrastructure. Unlike previous generations, in 3GSS some part of the image processing is distributed towards the sensor level by the use of intelligent cameras that are able to digitize and compress acquired analog image signals and perform image analysis algorithms like motion and face detection with the help of their attached digital computing components. The ultimate goal of 3GSS is to allow video data to be used for online alarm generation to assist human operators and for offline inspection effectively. In order to achieve this goal, 3GSS will provide smart systems that are able to generate real-time alarms defined on complex events and handle distributed storage and content-based retrieval of video data. The making of video surveillance systems smart requires fast, reliable and robust algorithms for moving object detection, classification, tracking and activity analysis. Starting from the 2GSS, a considerable amount of research has been devoted for the development of these intelligent algorithms. The given below figure show the process of video processing.

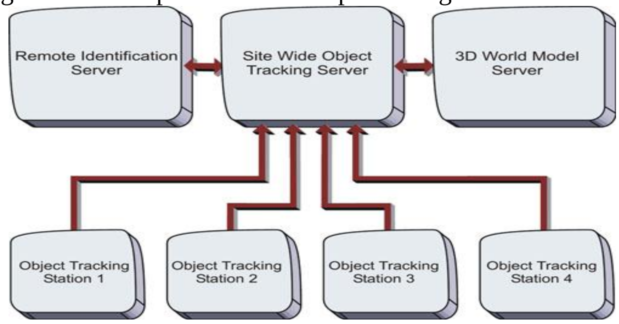

Fig 1.shows that online video object tacking for live video and movies

II. WORKING PROCESS OF VIDEO PROCESSING

Object Tracking stations Receive analogue or IP camera images directly and perform the most time consuming video content analysis and image processing tasks. OTS Stations implement the camera level and the scene level intelligent functions, like object tracking based on multiple camera views, shape classification, and detection of crossing perimeters of virtual zones. Results of the OTS calculations are collected and processed in a central Site Wide Object Tracking Server, which implements the site level intelligent functions, like identity tracking, evaluation of complex rules based on the identity and security clearance of moving persons or vehicles, and moreover this module is capable of recognizing the suspicious activities. The optional 3D World Model Server provides 3D calculation services for the other IDENTRACE modules. It can help OTS Stations in automatic calibration of the positions and viewing angles of cameras, while it can provide 3D location information to the SWOT Server, and the generation of synthesized virtual images for the central monitoring console. The Remote Identification Server helps the SWOT Server to handle the inevitable uncertainties of object tracking (e.g. when the system loses track of persons in blind areas, like rest rooms). In these situations the task is "just" to decide whether a person currently visible in the scene is or is not the same as the person who was seen before. So the RID Server should determine the identity of persons from a limited set of alternatives, which can be solved with a much higher reliability than those identification solutions have, which try to identify a person out of the 6 billion living on Earth. Moving object detection is the basic step for further

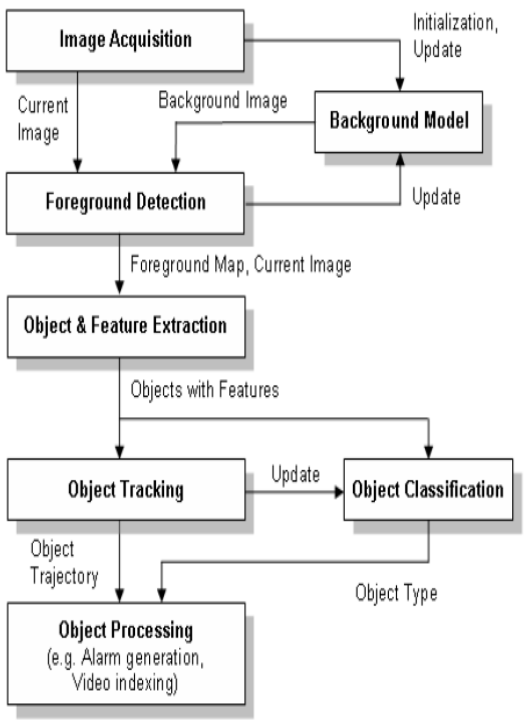

Fig.2. Process of video Pre-processing and Indexing.

analysis of video. It handles segmentation of moving objects from stationary background objects. This not only creates a focus of attention for higher level processing but also decreases computation time considerably. Commonly used techniques for object detection are background subtraction, statistical models, temporal differencing and optical flow. Due to dynamic environmental conditions such as illumination changes, shadows and waving tree branches in the wind object segmentation is a difficult and significant problem that needs to be handled well for a robust visual surveillance system. Object classification step categorizes detected objects into predefined classes such as human, vehicle, animal, clutter, etc. It is necessary to distinguish objects from each other in order to track and analyse their actions reliably. Currently, there are two major approaches towards moving object classification, which are shape-based and motion-based methods. The objects 2D spatial information whereas motion-based methods use temporal tracked features of objects for the classification solution. Detecting natural phenomenon such as fire and smoke may be incorporated into object classification components of the visual surveillance systems. Detecting fire and raising alarms make the human operators take precautions in a shorter time which would save properties, forests and animals from catastrophic consequences. The next step in the video analysis is tracking, which can be simply defined as the creation of temporal correspondence among detected objects from frame to frame. This procedure provides temporal identification of the segmented regions and generates cohesive information about the objects in the monitored area such as trajectory, speed and direction. The output produced by tracking step is generally used to support and enhance motion segmentation, object classification and higher level activity analysis. The final step of the smart video surveillance systems is to recognize the behaviours of objects and create high-level semantic descriptions of their actions. It may simply be considered as a classification problem of the temporal activity signals of the objects according to pre-labelled reference signals representing typical human actions. The outputs of these algorithms can be used both for providing the human operator with high level data to help him to make the decisions more accurately and in a shorter time and for offline indexing and searching stored video data effectively. The advances in the development of these algorithms would lead to breakthroughs in applications that use visual surveillance. Monitoring of banks, department stores, airports, museums, stations, private properties and parking lots for crime prevention and detection patrolling of highways and railways for accident detection. Measuring traffic flow, pedestrian congestion and athletic performance Compiling consumer demographics in shopping centre and amusement parks Extracting statistics from sport activities Counting endangered species Logging routine maintenance tasks at nuclear and industrial facilities Artistic performance evaluation and self-learning Law enforcement: Measuring speed of vehicles Detecting red light crossings and unnecessary lane occupation Military security: Patrolling national borders Measuring flow of refugees Monitoring peace treaties Providing secure regions around bases Detecting the natural phenomenon fire besides normal object motion would be an advantage of a visual surveillance system, thus, the presented system is able to detect fire in indoor and outdoor environments. Conventional point smoke and fire detectors typically detect the presence of certain particles generated by smoke and fire by ionization or photometry. An important weakness of point detectors is that they are distance limited and fail in open or large spaces. The strength of using video in fire detection is the ability to serve large and open spaces. Current fire and flame detection algorithms are based on the use of color and simple motion information in video [27]. In addition to detecting fire and flame colour moving regions, the method presented in this thesis analyses the motion patterns, the temporal periodicity and spatial variance of high-frequency. The given block diagram shows the process of video pre-processing model. Detecting regions that correspond to/ moving objects such as people and vehicles in video is the first basic step of almost every vision system since it provides a focus of attention and simples the processing on subsequent analysis steps. Due to dynamic changes in natural scenes such as sudden illumination and weather changes, repetitive motions that cause clutter (tree leaves moving in blowing wind), motion detection is a difficult problem to process reliably. Frequently used techniques for moving object detection are background subtraction, statistical methods, temporal differencing and optical flow whose descriptions are given below. Background subtraction is particularly a commonly used technique for motion segmentation in static scenes. It attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image that is created by averaging images over time in an initialization period. The pixels where the difference is above a threshold are classified as foreground.

III. OBJECT DETECTION AND TRACKING

The current system is able to distinguish transitory and stopped foreground objects from static background objects in dynamic scenes, detect and distinguish left and removed objects, classify detected objects into different groups such as human, human group and vehicle, track objects and generate trajectory information even in multi-occlusion cases and detect fire in video imagery. The computational complexity and even the constant factors of the algorithms we use are important for real time performance. Hence decisions on selecting the computer vision algorithms for various problems are affected by their computational run time performance as well as quality. Furthermore, the current system's use is limited only to stationary cameras and video inputs from Pan/Tilt/Zoom cameras where the view frustum may change arbitrarily are not supported. The system is initialized by feeding video imagery from a static camera monitoring a site. Most of the methods are able to work on both color and monochrome video imagery. The first step of our approach is distinguishing foreground objects.

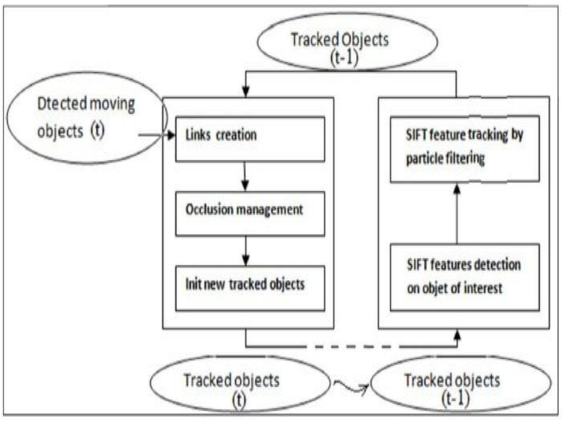

Fig 3: Process of object detection

Difficult problem. Almost the entire visual surveillance systems first step is detecting foreground objects. This both creates a focus of attention for higher processing levels such as tracking, classification and behavior understanding and reduces computation time Distinguishing foreground objects from the stationary background is both a significant and considerably since only pixels belonging to foreground objects need to be dealt with. Short and long term dynamic scene changes such as repetitive motions, light repentance, shadows, camera noise and sudden illumination variations make reliable and fast object detection Difficult. Hence, it is important to pay necessary attention to object detection step to have reliable, robust and fast visual surveillance system. The aim of object tracking is to establish a correspondence between objects or object parts in consecutive frames and to extract temporal information about objects such as trajectory, posture, speed and direction. Tracking detected objects frame by frame in video is a significant and difficult task. It is a crucial part of smart surveillance systems since without object tracking, the system could not extract cohesive temporal information about objects and higher level behavior analysis steps would not be possible. On the other hand, inaccurate foreground object segmentation due to shadows, reflectance and occlusions makes tracking a difficult problem. The information extracted by this level of tracking is adequate for most of the smart surveillance applications. Our approach makes use of the object features such as size, center of mass, bounding box and color histogram which are extracted in previous steps to establish a matching between objects in consecutive frames. Furthermore, our tracking algorithm detects object occlusion and distinguishes object identities after the split of occluded objects. By analyzing the object trajectory information, our tracking system is able to detect left and removed objects as well. Detecting left objects such as unattended luggage in airports or a car parked in front of a security sensitive building is important since these activities might be performed by terrorists to harm people. On the other hand, protecting objects against removal without permission has important applications such as in surveillance of museums, art galleries or even department stores to prevent theft. Due to these critical applications, left/removed object is important part of a surveillance system. Recent advances in multimedia compression technology, coupled with the significant increase in computer performance and the growth of Internet, have led to the widespread use and availability of digital video. Applications such as digital libraries, distance learning, video-on-demand, digital video broadcast, interactive TV, multimedia information systems generate and use large collections of video data. The main advantage of the clustering-based segmentation is that it is a generic technique that not only eliminates the need for threshold setting but also allows multiple features to be used simultaneously to improve the performance it involves analyzing intensity edges between consecutive frames. During a cut or a dissolve, new intensity edges appear far from the locations of the old edges. Similarly, old edges disappear far from the location of new edges. Thus, by counting the entering and exiting edge pixels, cuts, fades and dissolves are detected and classified. To obtain better results in case of object and camera movements, an algorithm for motion compensation is also included. It first estimates the global motion between frames that is then used to align the frames before detecting entering and exiting edge pixels. However, this technique is not able to handle multiple rapidly moving objects. Another weakness of the approach are the false positives due to the limitations of the edge detection method. In particular, rapid changes in the overall shot brightness, and very dark or very light frames, may cause false positives. The previous approaches for video segmentation process uncompressed video. As nowadays video is increasingly stored and moved in compressed format, it is highly desirable to develop methods that can operate directly on the encoded stream. Working in the compressed domain offers the following advantages. First, by not having to perform decoding/re-encoding, computational complexity is reduced and savings on decompression time and decompression storage are obtained. Second, operations are faster due to the lower data rate of compressed video. Last but not least, the encoded video stream already contains a rich set of pre-computed features, such as motion vectors (MVs) and block averages that are suitable for temporal video segmentation. Several algorithms for temporal video segmentation in the compressed domain have been reported.

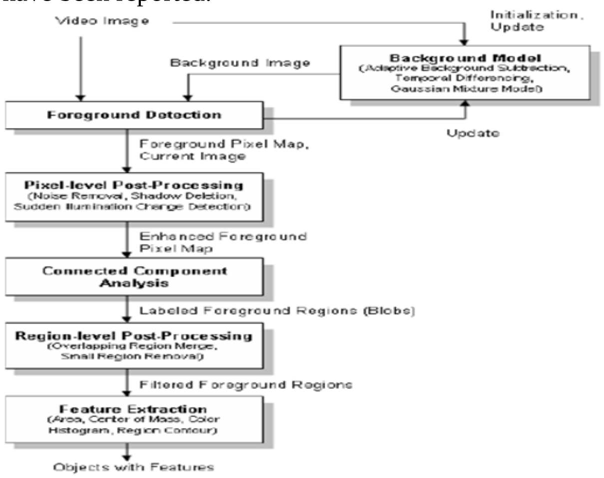

Fig 4: Object detection system

It is concerned with low-level visual processing and high-level image analysis, and is widely used in image understanding, human-computer interaction, surveillance, and robotics, to name a few. To tackle these challenges, this paper presents a tracking method that learns a robust object representation by partial least squares analysis and adapts to appearance change of the target and background while reducing drift. Many classes of objects can now be successfully detected with machine learning techniques. Face, cars, pedestrians and hands, has all been detected with low error rates by learning their appearance in a highly generic manner from extensive training sets. These recent advances have enabled the use of reliable object detection components in real systems, such as automatic face focusing functions on digital cameras. One key drawback of these methods, and the issue addressed here, is the prohibitive requirement that training sets contain thousands of manually annotated examples. We propose to reduce the requirement for such an extensive labelling by exploiting the temporal consistency occurring in a training video. The performance of this approach is evaluated on pedestrian detection in a surveillance camera setting, and on cell detection in microscopy data. This comes with virtually no loss in performance when compared to a standard learning procedure trained on a fully labelled sequence. In fact, in some cases, gains in performance are observed. Object detection and tracking is a major research area in computer vision. One of its application areas is traffic scene analysis

IV. PROBLEM FORMULATION

In the process of review we found that some performance affected problem related to the video object detection. These problem are affected the performance and accuracy of video object tracking and result overcome in fact of loss of frame. The segmentation region increase, decrease the accuracy and performance of object tracking. Some problems are mentioned here [4, 6].

1. Segmentation errors
2. Change of lighting conditions
3. Shadows
4. Occlusion
5. Automatic updating of background
6. False frame hit
7. Process of segmentation some frame value are lost.

V. MOTIVATION

The most intuitive way of segmenting and tracking objects is to first generate temporally- tracked homogeneous regions and then apply further processing to identify the semantic objects. For example, we may employ image segmentation algorithms, such as color or texture segmentation, to generate homogeneous regions in each frame and then track the motion of the segmented regions. We may also do it the other way around with motion segmentation methods. These methods typically first detect and estimate motion for each pixel and then group together that pixel s that has similar movement. For example, optical flow vectors may be used to estimate motion. Unfortunately, many difficulties arise from general videos to plague even the most sophisticated segmentation methods. Noise and clutter affect most pixel-based segmentation methods. While a multi-resolution algorithm can overcome the local in-formation restriction, it loses details at low-resolution levels and sometimes involves costly computation. Furthermore, occlusion and complex objects defy most non-model-based algorithms, but it is difficult to apply semantic models without higher-level features than pixels; hence, it is a chicken and egg problem. Typically, precise and accurate segmentation cannot be obtained for general videos. Without region information, it becomes non-trivial to extract multiple motions.

- Advantage of Video-based systems
  - Being able to capture a large variety of information
  - Relatively inexpensive
  - Easier to install, operate, and maintain
- Applications
  - Security surveillance
  - Home care surveillance
  - Intelligent transportation systems
- There is an urgent need for intelligent video systems to replace human operators to monitor the areas under surveillance.

VI. APPROACH USED FOR OBJECT TRACKING

The increasing rate of multimedia data and transmission facility induces some problem of data loss and delay of delivery. Now in the process of video object detection background updating is important factor for analysis. For the background updating used segmentation process and segmentation used clustering technique. Now in our dissertation used RBF neural network model for segmentation process and reduces the loss of frame and video data during object tracking process. The basic processing elements of neural networks are called artificial neurons, or simply neurons or nodes. In a simplified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the nonlinear characteristic exhibited by neurons is represented by a transfer function. The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm. The basic architecture consists of three types of neuron layers: input, hidden, and output layers. In feed-forward networks, the signal flow is from input to output units, strictly in a feed-forward direction. The data processing can extend over multiple (layers of) units, but no feedback connections are present. Recurrent networks contain feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such that the network will evolve to a stable state in which these activations do not change anymore.

VII. CONCLUSION

Object tracking and segmentation has most importance in video processing systems. This paper focuses on the major issue of segmented area and frame loss in video object tracking & segmentation and their detection techniques. Researchers developed several object tracking & segmentation methods but there is not any technique which can have highest segmented area and least error. For increasing segmented area and reducing frame loss one needs special technique except existing ones because of their respective disadvantages. Therefore, designing such type of security technique is still an open research challenge.

VIII. ACKNOWLEDGMENT

I express my sincere gratitude to Prof. Manoj Yadav, Head Department of CSE, for his valuable guidance and advice. Also I would like to thanks to my guide Prof. Parmalik Kumar and the faculty members for their continuous support and encouragement.

REFERENCES

[1] B. Babenko, M.-H. Yang, and S. Belongie, “Visual tracking with online multiple instances learning” IEEE Comput. Vis. Pattern Recognit. Conf., San Diego, CA, Jun. 2009, pp. 983–990.

[2] R. T. Collins, Y. Liu, and M. Leordeanu, “Online selection of discrimina-tive tracking features” IEEE Trans. Pattern Anal. Mach. Intell., vol. 27,no. 10, Oct. 2005, pp. 1631–1643.

[3] M. Everingham, L. V. Gool, C. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (VOC) challenge”Int. J. Comput. Vis.,vol. 88, no. 2, 2010 pp. 303–338.

[4] H. Grabner, C. Leistner, and H. Bischof, “Semi-supervised on-line boosting for robust tracking” inProc. Eur. Comput. Vis. Conf., 2008, pp. 234–247.

[5] Z. Kalal, J. Matas, and K. Mikolajczyk, “P-N learning: Bootstrapping binary classifiers by structural constraints” inProc. IEEE Comput. Vis. Pattern Recognit. Conf., San Francisco, CA, Jun. 2010, pp. 49–56.

[6] J. Kwon and K. Lee, “Visual tracking decomposition” in Proc. IEEE Comput. Vis. Pattern Recognit. Conf., San Francisco, CA, Jun. 2010, pp. 1269–1276.

[7] X. Li, W. Hu, Z. Zhang, X. Zhang, and G. Luo, “Robust visual tracking based on incremental tensor subspace learning” inProc. IEEE Comput. Vis. Pattern Recognit. Conf., Rio de Janeiro, Brazil, Oct. 2007

[8] D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, “Incremental learning for robust visual tracking” Int. J. Comput. Vis., vol. 77, nos. 1–3, pp. 2008 125–141.

[9] Zhiding Yu, Chunjing Xu, Jianzhuang Liu, “Nonparametric density estimation on a graph: Learning framework, fast approximation and

application in image segmentation" Oscar C 2011 Association of Computing Machinery, Scottsdale, USA, Nov 2011.

[10] M Zaki, M Youssef. "A comparative simulation study of link quality estimators in wireless sensor networks", SIFNAL, IMAGE AND VIDEO PROCESSING, London, vol. 3, 2009.

[11] Gary Bradski and Adrian Kaehler. Learning OpenCV (Adrian Kaehler O'Reilly Media, Inc). 2008.

[12] Yang, Y., Hallman, S., Ramanan, D., & Fowlkes, C."Layered object detection for multi-class segmentation" InIEEE conference on computer vision and pattern recognition 2010.

[13] Yu, H. F., Hsieh, C. J., Chang, K. W., & Lin, C. J. "Large linear classification when data cannot fit in memory" In ACM SIGKDD conference on knowledge discovery and data mining 2010.

[14] Van de Sande, K. E. A., Gevers, T., & Snoek, C. G. M."Evaluating color descriptors for object and scene recognition" IEEE Transactions on Pattern Analysis and Machine Intelligence, 9,1582–1596,2010.

[15] Vedaldi, A., & Zisserman, A." Efficient additive kernels via explicit feature maps" In IEEE conference on computer vision and pattern recognition 2010 IEEE.

[16] Vedaldi, A., Gulshan, V., Varma, M., & Zisserman, A. "Mul-tiple kernels for object detection"In International conference on computer vision 2009.

[17] Maire, M., Arbelaez, P., Fowlkes, C., & Malik, J."Using con-tours to detect and localize junctions in natural images". In 2008 IEEE.

[18] Malisiewicz, T.,& Efros, A."Improving spatial support for objects via multiple segmentations".conference on computer vision and patter n recognition. In British machine vision conference 2007.

[19] Malisiewicz, T., & Efros, A. A."Recognition by association via learning perexemplar distances". In IEEE conference on computer vision and pattern recognition 2008 IEEE.

[20] F. Tang, S. Brennan, Q. Zhao, and H. Tao, "Co-tracking using semisupervised support vector machines" in Proc. of the International Conference on Computer Vision (ICCV). Rio de Janeiro, Brazil: IEEE Computer Society Washington, DC, USA, 2007, pp. 1–8.

[21] Qing Wang, Feng Chen, Wenli Xu, and Ming-Hsuan Yang "Object Tracking via Partial LeastSquares Analysis" 1057–7149/2012 IEEE.

[22] Yi We i Zhao Long "Robust objects tracking algorithm based on adaptive background updating" 978-1-4673-0311-8/12/2012 IEEE.

[23] João Carreira·Fuxin Li·Cristian Sminchisescu "Object Recognition by Sequential Figure-Ground Ranking" Int J Comput Vis DOI 10.1007/s11263-011-0507-2.2011.

[24] Samuele Salti, Andrea Cavallaro, Luigi Di Stefano "Adaptive Appearance Modeling for Video Tracking:Survey and Evaluation" ACCEPTED FOR PUBLICATION IN IEEE TRANSACTIONS ON IMAGE PROCESSING 2011.

[25] Karim Ali,David Hasler,Franc¸ois Fleuret "FlowBoost Appearance Learning from Sparsely Annotated Video" 2011 IEEE.

[26] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas" Tracking-Learning-Detection" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 6, NO. 1, 2010.

[27] Zdenek Kalal, Jiri Matas, Krystian Mikolajczyk "P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints" Published at the 23rd IEEE Conference on Computer Vision and Pattern Recognition, CVPR, June 13-18, San Francisco, CA, USA, 2010.

[28] Severin Stalder , Helmut Grabner , and Luc Van Gool "Cascaded Confidence Filtering for Improved Tracking-by-Detection" 2011 IEEE.

[29] Michael D. Breitenstein, Fabian Reichlin,Bastian Leibe,Esther Koller-Meier,Luc Van Gool "Robust Tracking-by-Detection using a Detector Confidence Particle Filter" 2009 IEEE 12th International Conference on Computer Vision (ICCV) 978-1-4244-4419-9/2009 IEEE.

[30] ZuWhan Kim "Real Time Object Tracking based on Dynamic Feature Grouping with Background Subtraction" 978-1-4244-2243-2/08 2008 IEEE.

[31] P. KaewTraKulPong and R. Bowden "An Improved Adaptive Background Mixture Model for Real time Tracking with Shadow Detection" In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, AVBS01. Sept 2001.

[32] Irena Koprinska,, Sergio Carrato "Temporal Video Segmentation: A Survey" Image Processing Laboratory, Universi t y of Trieste via Valeri o 10, 34127 Tri este, ITALY.

