International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 2, March - April 2017 ISSN 2278-6856

A Survey of Intelligent Surveillance Systems

Dr. Abdulamir A. Karim[1], Narjis Mezaal Shati[2]

[1]University of Technology/ Department of Computer Science Baghdad, Iraq

[2]Al-mustansiryah University / College of Science Baghdad, Iraq

Abstract

Recent technology and market trends have demanded the need for solutions to video/camera systems and analytics. This paper highlights the video system architectures, tasks, and related analytic methods. It demonstrates that the importance of the role that intelligent video systems and analytics play can be found in a variety of application such as homeland security, crime prevention, traffic control, accident prediction and detection, and monitoring patients, elderly and children at home, etc. Research directions are outlined in the end with a focus on what is essential to achieve the goals of intelligent video systems and analytics.

Keywords: Intelligent surveillance system, Smart surveillance system, Automatic video surveillance, Video analytic**s,** Video content analysis.

1. INTRODUCTION

Intelligent video surveillance of dynamic and complex scenes is one of the most active research topics in Computer Vision. It aims is to efficiently extract useful information from a huge amount of videos collected by surveillance cameras by automatically detecting, tracking and recognizing objects of interest, understanding and analyzing their activities from image sequences [1,2]. Given the tremendous number of video data produced every day, there is a great demand for automated systems that analyze and understand the events in these videos. Retrieving and identifying human activities in videos has become more interesting due to its potential applications in real life. Including: automated video surveillance systems, human-computer interaction, assisted living environments and nursing care institutions, sports interpretation, video annotation and indexing, and video summarization [3].
The progression of surveillance systems can be decomposed into three generations [4].

1.1 First-Generation Video Surveillance Systems (1GSS)

This is considered the beginning of modern surveillance systems prevailed from the 1960s to the 1980s. Video cameras were exploited and installed to procure visual signals from remote locations and subsequently viewed from the control room. The human administration in the control rooms would view a large quantity of monitors and attempt to espy events of interest. The main di□culties in 1GSS is the natural human attention span, which resulted in a prominent amount of missed events. Technical di□culties consisted of bandwidth requirements of the sensors, the susceptibility to noise and degradation of contemporary playback mechanisms, and the storage of video surveillance tapes. The interrelationships of 1GSS are presented in Figure 1 [4].

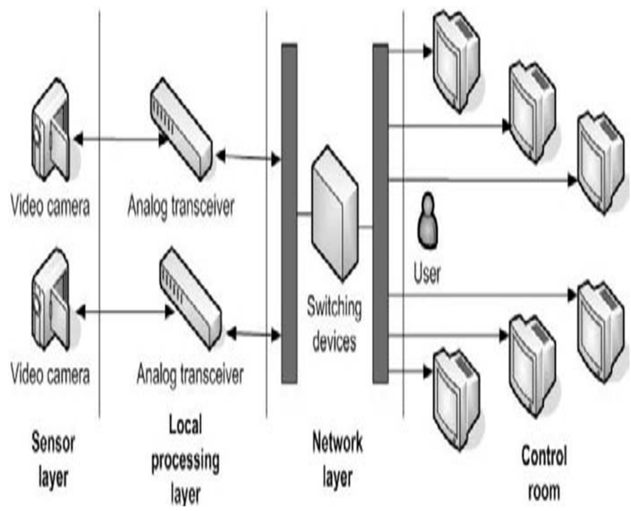

Figure 1: "1GSS" Structural example [4].

1.2 2nd-G Surveillance Systems (2GSS)

Between duration (1980-2000) is the period of 2GSS. In the 80s, video processing and event detection improved according to the development of video cameras, the low cost computers, Also development in communication systems that provide higher quality in low expense. Digital technologies that used present "digital compression, strong dispatch, decrease bandwidth, and techniques". This lead to memorable enhancement in outcomes encompass "analysis in real-time and tracking capabilities, individual identifying and understanding its behavior, data fusion of multi-sensor, wired/ wireless networks, and high-rendering algorithms" that making easy the duties of the administrators of the "control rooms" [4].

1.3 3rd-G Surveillance Systems (3GSS)

Onset in 2000, 3GSS finish the digital conversion' and it is became completely employs digital information starting in low-level by using sensors to high-level analysis and outcomes. This due to the increasing in networks processing speeds, and the decreasing expenses in utilization of communications, and reaching a new level in the heterogeneity and mobility of sensors. Figure 2 show 3GSS connections [4].

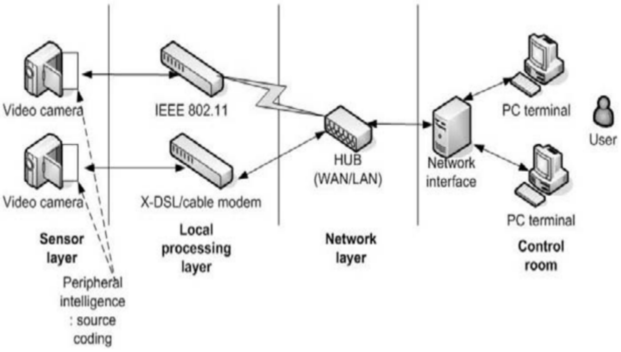

Figure 2: "3GSS" Structural example [4].

On future generations, technological evolution will continue to have a senior effectiveness due to development in to all the different fields, hardware, software, and the communication.

2. VIDEO SURVEILLANCE SYSTEM CLASSES

The classification of video surveillance systems performed according to key features such as (Background nature, Number of objects to track, Size of the monitored area, and Evaluation method) [1].

2.1 Background nature

It is concerning the properties of the environment to be monitored. Can be static (or nearly static) or dynamic depending on the environment we are observing.

- A static background can be a lab or an office, where the environment is mostly static, the light is artificial and the 3D structure of the environment is known [5, 6, 7]. An outdoor example of a nearly static background is a highway: the direction of the car flow is known and all the cars travel in separate lanes [8]. Another typical outdoor static background scenario is a parking lot [1].
- An example of dynamic background environment can be a water scenario because of the rays of sun on the water surface and waves made by moving steamships or by wind that form correlated moving patterns [9, 10, 11]. Similar problems arise in background such as trees and lawns with wind and a great amount of shadow [12]. A crowded environment with furniture that is moved continuously (emergency room, in airport, in Malls) represent an indoor dynamic background [13]. Figure 3 shows different background examples.

2.2 Number of objects to track

Identify Number of tracked objects in order to manage crowded or non-crowded situations. The number of objects to track is a key aspect for classifying a system. Tracking isolated objects is not a difficult task. Failures arise when the tracking system has to deal with occlusions and multiple objects close to each other. Usually up to 3 or 4 objects (e.g., people) are considered in the scene at the same time [5, 6]. Dealing with more objects is challenging because of partial and complete occlusions causing tracking failures. Taking into account more than 10 objects in the scene considered a hard task [14].

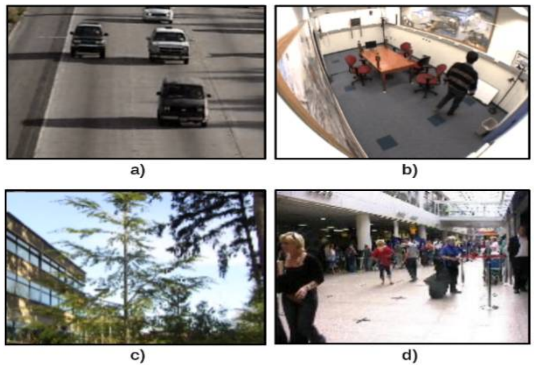

Figure 3: Different background examples: a) static outdoor [8], b) static indoor [6], c) dynamic outdoor [12], d) dynamic indoor [13].

2.3 Size of the monitored area

It is concerning number, position and type of installed cameras. To monitor an entrance or a corridor a single camera system can be sufficient, but using multiple cameras for oversight be wide areas or/and managing occlusions. Generalization of a system with single-camera to be a multi-camera system this arise several problems such as "installation and calibration of camera, object matching, and data fusion" [15].

A "blind spots" caused by a shortage of cameras, but installing redundant cameras raise time of processing, complexity of algorithmic, cost of installation, complexity of calibration, and complexity of object matching (finding correspondences between objects in different scene) [1].
Different types of cameras used for grabbing images. Using a stereo camera gives the chance to obtain 3D data useful in solving partial occlusions. Exploiting a stereo camera needs calibration and it costs twice as a single camera [16].

3. VIDEO SURVEILLANCE TASK

Automatic video surveillance task can be broken down into a series of sub problems [17], as shown in figure (4).

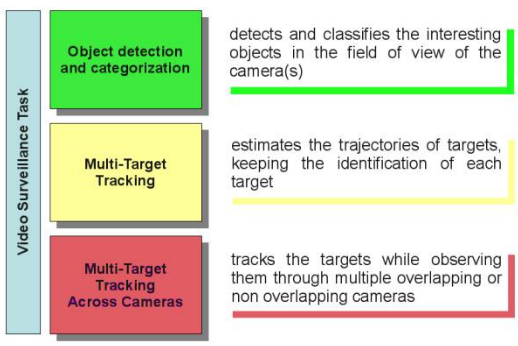

Figure 4: Automatic video surveillance task [17].

Furthermore, the main task in ISS is Activity analysis; it direct activities into different classes and find out normal and abnormal activities. Intelligent surveillance system should be able to react to particular events (e.g., an alarm could be send if an unauthorized vehicle enters a restricted area) [18]. A multi-camera video surveillance data flow schema shown in figure 5.

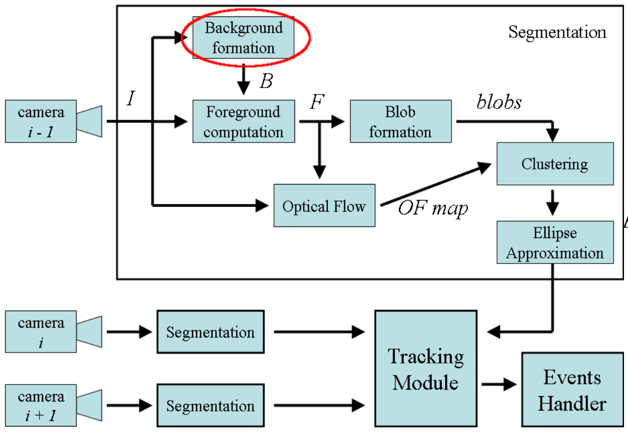

Figure 5: A multi-camera video surveillance data flow schema. Background modelling module is circled [17].

3.1 Object Detection

In a VSS (video surveillance system), the goal is to detach regions belonging to the background from the regions of the foreground (dynamic objects). The common used image segmentation approaches with dynamic background categorized as follows:

3.1.1. Temporal differencing

It is the most basic, but efficient approach. Adjacent "Frame Difference" works by subtract current pixel value from its previous one, tagged it as foreground if a threshold is less than an absolute difference [19]. With high frequencies elements (such as waves, trees, etc...) for calculating background, this is the worse approach, due to the impossibility of taking into account periodical and multimodal changes in the background. While it is very adaptive to animated environments, it does a needy job of elicitation all the relevant pixels, (having punctures left inside animated objects) [20].

3.1.2 Optical flow

It is an approximation of the local image motion based on local derivatives in a sequence of images. [21].
In image sequence to detect animating areas, optical flow based motion segmentation utilizes features of flow-vector of animated objects during the time. Methods based on optical flow used to separate animated objects. Drawback of these approaches are hypersensitive to noise, complicated, and in real time not used without particularist hardware [15].

3.1.3 Background subtraction.

Background subtraction (BS) is a widely used segmentation technicality performed in real-time; in which data from each camera are processed through two major steps: Background modelling, Foreground extraction [22], as figure (6) shown.

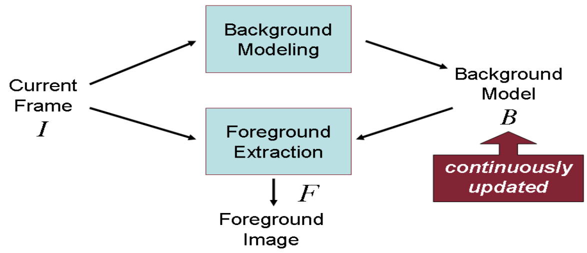

Figure 6: Background subtraction [17].

A. **Background modelling:**

The carried out of this process is creating a model that represents the regions of the scene that remain constant in time. The model must be updated over time to take the disparity of environment conditions, like changes in illumination (sudden and gradual), shadows, jitter in camera, background elements movement (waves on water surfaces, swaying of trees), and the geometry of background changes (such as car-park). For background modelling, and updating the model a different methods have been used. Statistical models (single Gaussians/mixture of Gaussians) are used widely, other models such us median, minimum-maximum values, etc. are commonly used [23,24,25,26].

A classification of background subtraction methods by identifying two classes, namely (recursive & non-recursive) techniques. In recursive methods with each new video frame update the used single background model. In non-recursive methods utilized a buffer L of n previous frames and from its statistical properties guess the background model [27, 28].

Another classification (predictive / non-predictive). Predictive approaches shape the scene as a time series and evolve an animated model that based on past observations to recover the current input. Non-predictive use the observations at a specific pixel to construct a probabilistic representation (pdf) with neglecting the order of the input observations [29]. Table (1) shows the classes of background modeling.

Table1: Background modeling methods.

Buffered Based	Predictive
Median Filter	Kalman Filter
Minimum Maximum Filter	Weiner Filter
Mediod Filter	Hidden Markov Models
Eigenbackgrounds	

Recursive	Non-Parametric
Approximated Median Filter	Kernel Density Estimation
Single Gaussian	Condensation
Mixture of Gaussians	
Adaptive Mixture of Gaussians	

B. Foreground extraction:

After computing the background model and the distinction between the current frame and this model calculated, the animated objects (foreground) detect. This operation output is a binary mask called foreground image containing the moving objects [9], as seen in figure (7).
Analyzing the binary image in order to find connected components (blobs) representing the moving objects silhouettes then perform segmentation [9].

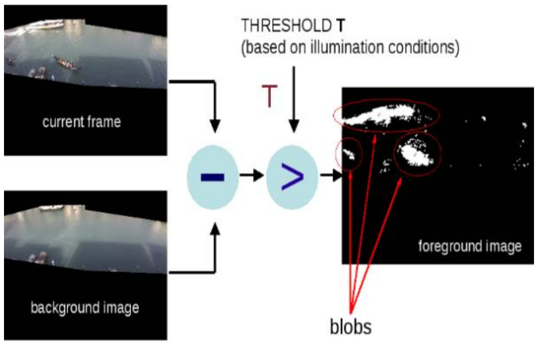

Figure 7: Background subtraction technique [9].

3.2 Tracking Object of Interest

Multi-Target Tracking (MTT) involves determining correspondences between a set of observations (or measurements) separated over time. The general MTT problem concerns with multiple targets and multiple measurements, therefore each target needs to be validated and associated to a single measurement in a data association process. MTT is a challenging task when partial and complete occlusions occur among the targets (e.g., crowded environments) [30, 31]. Figure (8) illustrate the problems of MTT.

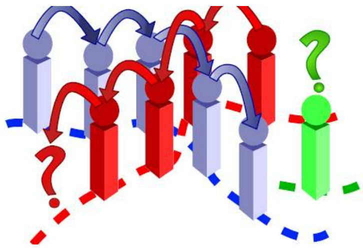

Figure 8: Multitarget tracking (MTT) and localization [31].

The MTT process split to two phases:

- ***association***: assigning each forthcoming observation to specified object-track.
- ***estimate***: for associated track, supply state estimate using the observation that received.

At every receive of fresh observation, it assigned to accurate track among set of present tracks, and create a new track if the track is not existing in the set of present tracks. Some techniques of association of data and management of tracks needs for this tracking system [32].
In a single hypothesis approach, assigning the observation to one of the existent tracks at all times and the track management module has to deal with initialization of track, update of track (containing prediction & data association), and deletion of track. The major difficulty of such an approach is that the system can't recover from a wrong association (observation associated to faulty track). Since in a crowded environment isn't simple to specify an observation to a particular track, it is advisable to use a multi-hypothesis tracking system, figure 9 illustrate the tasks of track managements [32].

Track initialization	When a new observation is obtained, if it is not highly correlated with any existing track, a new track is created.
Track update	Once observations are associated to tracks in a one-to-one fashion, standard updates of the 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.
Track merge	If the probability association between two tracks exceeds a threshold, one of the two tracks is deleted, keeping only the most significant hypothesis.
Track deletion	When a track is not supported by observations, the uncertainty in the state estimate increases. If the uncertainty reaches a threshold, the track is deleted.

Figure (9): Tracks management tasks [32].

In a multi-hypothesis approach, assigned the observation to more than one track, the system split every candidate tracks into two new ones (one updated with the observation, one not updated). Such a technique (split of track) leads to a possible increasing the number of tracks, so the system requires detect and delete redundant ones (track merging) [32]. Figure 10 illustrate visual tracking methods classification.

4. DATA FUSION

The general data fusion problem concerns with merging the data of multiple sensors (cameras), in order to obtain accurate information, this improved accuracy from existing sensors [33, 34]. In numerous domains, such as target tracking in military or autonomous robotics, the multi Sensor Data Fusion (MSDF) used [35].

Even if segmentation and tracking processes can be carried out by single-camera systems, wide areas can be covered only through multiple-camera systems. Moreover, even if dealing with a small area, multiple views can address the problem of handling occlusions and can provide 3D information about the detected objects. The use of multiple cameras lead to series of problems (installation of camera, matching of object, and data fusion) [36].

Data coming from multiple tracking modules processed in order to obtain a unique data representation. Usually each tracking module outputs data that are in a local coordinate system: the data fusion module aims at transforming all the data in a global coordinate system [37]. Multi-camera data fusion example illustrated in figure 11.

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 2, March - April 2017 ISSN 2278-6856

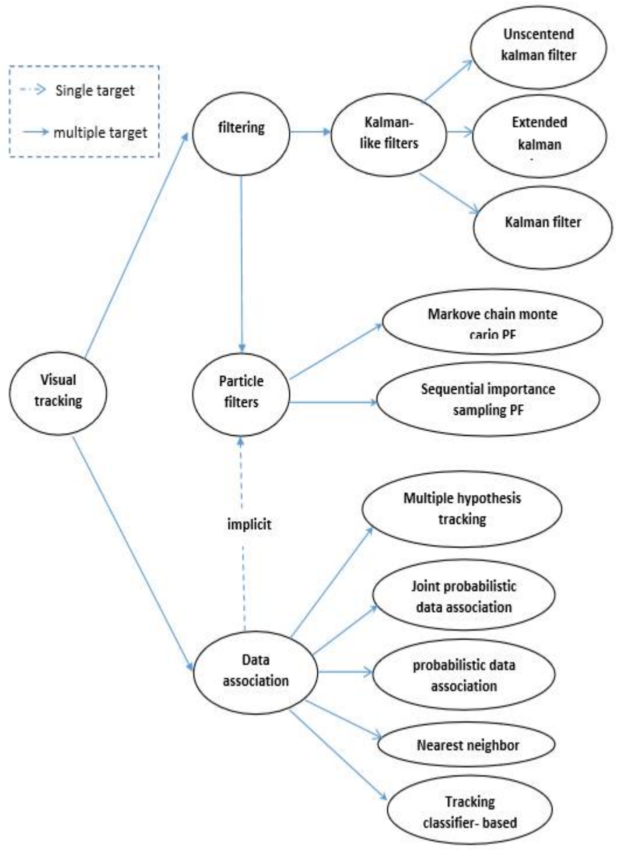

Figure 10: Visual tracking methods [32].

Figure 11: Multi-camera data fusion example: three camera frames are rectified and stitched into a composite image [37].

5. Evaluation Method

Carrying on a series of experiments is a key requirement to evaluate the accuracy of a system. It is possible to use:

- Self-recorded data for performing a self – evaluation. Researchers to test their algorithms use this data. The system evaluation cannot be compared with previous works because this data are not publicly available.
- Publicly available benchmark dataset such as PETS [38], VISOR [39], ATON [40], WALLFLOWER [12], and CAVIAR [42] for computing quantitative results. It is not always possible to use benchmarks (as an example, benchmarks for video surveillance using stereo cameras do not exist) or to obtain a third party evaluation.

Using *benchmark data*, which give the chance to quantitatively evaluate the performances of the method with respect to others.

- Third party evaluation for obtaining an objective certification.

A third party evaluation provide an objective evaluation of the system, but it is not always possible to perform, according to the requirement of developing a user interface to allow the test performances.

6. Conclusions

In this paper summarization of the development of intelligent surveillance systems and listed the representative works for researchers to have general overview of the state, so as to evolve an intelligent video surveillance system with the following features:

- ✓ Able to model dynamic types of background,
- ✓ Able to track a great number of objects,
- ✓ That use multiple cameras for covering large areas and handling occlusions,
- ✓ That can be valued using objective criteria.

References

[1] Dee, H. M.; and Velastin, S. A.; "How close are we to solving the problem of automated visual surveillance? A review of real-world surveillance, scientific progress and evaluative mechanisms"; Machine Vision and Applications: Special Issue on Video Surveillance Research in Industry and Academic; 2007.

[2] Qian, H.; Wu, X.; and Xu, Y.; "Intelligent Surveillance Systems"; Springer Publishing Company, Inc.; 2011.

[3] Roshtkhari, M. J.; "Visual Event Description in Videos"; PhD. Thesis; McGill University; Faculty of Graduate Studies and Research; Department of Electrical and Computer Engineering; Montreal; 2014.

[4] Pal, S. K.; Petrosino, E.; and Maddalena L.; "Handbook of Soft Computing for Video Surveillance"; CRC press; 2012.

[5] Tsai, Y.T.; Shih, H. C.; and Huang, C. L.; " Multiple human objects tracking in crowded scenes "; In ICPR'06: Proceedings of the 18th International Conference on Pattern Recognition; PP. 51-54; 2006.

[6] Gilbert, A.; and Bowden, R.; "Multi person tracking within crowded scenes"; In Workshop on Human Motion; PP. 166-179; 2007.

[7] Leo, M.; and Et al.; "Automatic monitoring of forbidden areas to prevent illegal accesses"; In ICAPR (2); PP. 635-643; 2005.

[8] Marchetti, L.; Nobili, D.; and Iocchi, L.; "Improving tracking by integrating reliability of multiple sources"; In 11th International Conference on Information Fusion; PP. 1-8; 2008.

[9] Ablavsky, V.; "Background models for tracking objects in water"; In ICIP (3); PP. 125-128; 2003.
[10] Monnet, A.; and Et al.; "Background modeling and subtraction of dynamic scenes"; In ICCV; PP. 1305-1312; 2003.
[11] Spencer, L.; and Shah, M.; " Water video analysis "; In ICIP; PP. 2705-2708; 2004.
[12] WALLFLOWER; Test images for wallflower paper; URL: http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/ testimages.htm.
[13] PETS2007; dataset for Performance evaluation of tracking systems 2007; URL: http://www.pets2009.net.
[14] Bloisi, D. D.; and Et al.; "A novel segmentation method for crowded scenes"; In Proc. of 4th International Conference on Computer Vision Theory and Applications; 2009.
[15] Hu, W.; and Et al.; "A survey on visual surveillance of object motion and behaviors"; In IEEE Transactions on Systems, Man, and Cybernetics; PP. 334-352; 2004.
[16] Iocchi, L.; and Bolles, R. C.; "Integrating plan-view tracking and color-based person models for multiple people tracking"; In Proc. of the International Conference on Image Processing; PP. 872-875; 2005.
[17] Javed, O.; and Shah, M.; " Automated Multi-Camera Surveillance: Algorithms and Practice "; Springer Publishing Company, Inc.; 2008.
[18] Liu, H.; Chen, S.; and Kubota N.; "Intelligent Video Systems and analytics: Survey"; In IEEE; 2013.
[19] Toyama, K.; and Et al.; "principles and practice of background maintenance "; In Proc. of the 7th IEEE International Conference on Computer Vision; Vol. 1; PP. 255-261; 1999.
[20] Cheng, L.; and Et al.; "An online discriminative approach to background subtraction"; In IEEE International Conference on Advanced Video and Signal based Surveillance; pp. 2; 2006.
[21] Barron, J. L.; and Thacker, N. A.; " Tutorial: Computing 2D and 3D Optical Flow"; Tina Memo; No. 012; 2004.
[22] Benezeth, Y.; and Et al.; "Review and evaluation of commonly-implemented background subtraction algorithms"; In ICPR; pp. 1-4; 2008.
[23] Wren, C. R.; and Et al.; " Realtime tracking of the human body "; In IEEE Transactions on Pattern Analysis and Machine Intelligence; vol. 19, no. 7, pp. 780-785; 1997.
[24] Stau□er, C.; and Grimson, W. E. L; "Adaptive background mixture models for real-time tracking"; In IEEE Computer Society Conference on Computer Vision and Pattern Recognition; vol. 2, pp. 246-252; 1999.
[25] Cucchiara, R.; and Et al.; " Detecting moving objects, ghosts, and shadows in video streams "; In IEEE Transactions on Pattern Analysis and Machine Intelligence; vol. 25, no. 10, pp. 1337-1342; 2003.
[26] Haritaoglu, I.; Harwood, D.; and Davis, L.S.; " W4: A real time system for detecting and tracking people "; In IEEE Computer Society Conference on Computer Vision and Pattern Recognition; vol. 0, pp. 962; 1998.
[27] Cheung, S.; and Kamath, C.; "Robust techniques for background subtraction in urban tra□c video "; In Visual Communications and Image Processing 2004; vol. 5308, pp 881-892; 2004.
[28] Parks, D.; and Fels, S.; " Evaluation of background subtraction algorithms with post-processing "; In Proceedings of the 2008 IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance; pp. 192-199; 2008.
[29] Mittal, A.; and Paragios, N.; " Motion-based background subtraction using adaptive kernel density estimation "; In Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; pp. 302-309; 2004.
[30] Shalom, Y. B.; "Tracking and data association"; Academic Press Professional, Inc.; 1987.
[31] Laet, D. L; "Rigorously Bayesian Multitarget Tracking and Localization"; PhD. Thesis; Katholieke University Leuven; Faculty of Engineering; Belgium.
[32] Hall, D.; and Et al.; "Comparison of target detection algorithms using adaptive background models "; In Proc. of the 14th International Conference on Computer Communications and Networks; pp. 113-120; 2005.
[33] Chong, C. Y.; and Et al.; "Architectures and algorithms for track association and fusion."; In In IEEE Aerospace and Electronic Systems; vol. 15, pp. 5-13; 2000.
[34] Hall, D. L.; and Llinas, J.; " Handbook of Multisensor Data Fusion "; CRC Press; 2001.
[35] Luo, R. C.; Yih, C. C.; and Su, K. L.; "Multisensor fusion and integration: approaches, applications, and future research directions "; In Sensors Journal IEEE; vol. 2, pp. 107-119; 2002.
[36] Yilmaz, A.; Javed, O.; and Shah M.; " Object tracking: A survey "; In ACM Comput. Surv.; vol. 38, no. 4, pp. 1-45; 2006.
[37] Liebowitz, D.; and Zisserman, A.; " Metric rectification for perspective images ofplanes "; In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; pp. 482-488; 1998.
[38] PETS2009; dataset for Performance evaluation of tracking systems 2009; URL: http://pets2007.net.
[39] VISOR; Video surveillance online repository; URL: http://imagelab.ing.unimore.it/visor.
[40] ATON; Autonomous agents for on-scene networked incident management; URL: http://cvrr.ucsd.edu/aton/testbed.
[41] CAVIAR; Context aware vision using image-based active recognition; URL: http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1