COMMUNICATIONS IN SCIENCE AND TECHNOLOGY
Homepage: cst.kipmi.or.id

A comprehensive review on intelligent surveillance systems

Sutrisno Ibrahim*

Electrical Engineering Department, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia

Article history:
Received: 28 April 2016 / Received in revised form: 11 May 2016 / Accepted: 17 May 2016

Abstract

Intelligent surveillance system (ISS) has received growing attention due to the increasing demand on security and safety. ISS is able to automatically analyze image, video, audio or other type of surveillance data without or with limited human intervention. The recent developments in sensor devices, computer vision, and machine learning have an important role in enabling such intelligent system. This paper aims to provide general overview of intelligent surveillance system and discuss some possible sensor modalities and their fusion scenarios such as visible camera (CCTV), infrared camera, thermal camera and radar. This paper also discusses main processing steps in ISS: background-foreground segmentation, object detection and classification, tracking, and behavioral analysis.

Keywords: Intelligent surveillance system (ISS), object detection, human detection, moving object detection, object tracking, object recognition, behavioral analysis, CCTV.

1. Introduction

Massive amount of security cameras, along with other sensors, have been deployed to monitor critical infrastructure such as: military bases, airport, power plant, banking, campuses, etc. Manual monitoring by human operator is inefficient solution or even unpractical because human resource is expensive and has limited ability [1]. Intelligent surveillance system (ISS) is envisioned to automatically monitor the environment or infrastructure with less or without human intervention. Such monitoring tasks include automatically detecting and tracking object (like human or vehicle) and performing further analysis and actions. Signal processing, image processing, and artificial intelligence (machine learning) techniques play important role to develop such intelligent system.

Visible camera such as CCTV is the most common modalities (device) for surveillance system. It has long been in use to monitor environments, people, events and activities. Extensive studies have been conducted to automatically analyze data (image or video) from surveillance camera. Much of these studies have been discussed in several focused review papers: background-foreground segmentation [2-7], objects detection and classification [8-10], tracking [11-14], and behavioral analysis [15-17]. Different sensor modalities other than visible camera have been explored also for surveillance system such as infrared camera and thermal camera [18], radar (radio detection and ranging) [19, 20], lidar (light detection and ranging) [21-23], audio sensor [24], etc. Several review papers have discussed also different techniques for sensor fusion [25-29] to improve the system performance. However, there is still lack of comprehensive paper that discusses general overview of intelligent surveillance system.

The main objective of this paper is to provide general overview of intelligent surveillance system and review the existing methods for each its processing steps. The rest of this paper is organized as follows: Section II presents an overview of intelligent surveillance systems. Followed by Section III discusses some possible sensor modalities and different fusion scenarios. Section IV reviews the existing methods for background-foreground segmentation, object detection, classification, tracking, and behavioral analysis. Section V concludes the paper and highlights future research direction in this field.

2. Intelligent surveillance system (ISS) overview

Huge amount of security cameras, intelligent surveillance system (ISS) is a surveillance system that has intelligent capability to automatically analyze surveillance data and perform necessary actions such as generating alarm or warning. ISS is interdisciplinary topic that involves electronic (sensing device), computer vision and pattern recognition, artificial intelligence (machine learning), networking, communication and other areas. Intelligent surveillance system is promising to be implemented in various environments and applications. Some typical applications are listed as follows.

* Corresponding author.
Email: suibrahim@ksu.edu.sa

© 2016 KIPMI

- Home security [30, 31] and intrusion detection [32].
- Home care and safety [33].
- Public transport area such as airport, seaport, bus/train terminal [24].
- Public area [34] such as colleges, campuses, governmental building.
- Traffic monitoring [35].
- Crown management and analysis [36].
- Pedestrian detection and autonomous car [9, 10].
- Remote military surveillance, border monitoring, perimeter surveillance for power plant, company, etc.

Example of surveillance systems that have been previously studied or developed to have automation or intelligent capabilities: VSAM (video surveillance and monitoring) [37], W4 [38], PRISMATICA (pro-active integrated systems for security management by technological institutional and communication assistance) [24, 39], ADVISOR (annotated digital video for intelligent surveillance and optimized retrieval) [40]. Fig. 1 shows the overview of PRISMATICA system that has been proposed to improve passenger security and safety in the public transport system. It contains several main components: camera network (existing CCTV), intelligent camera system, transmission system, audio surveillance, operator, and also the main server (MIFSA).

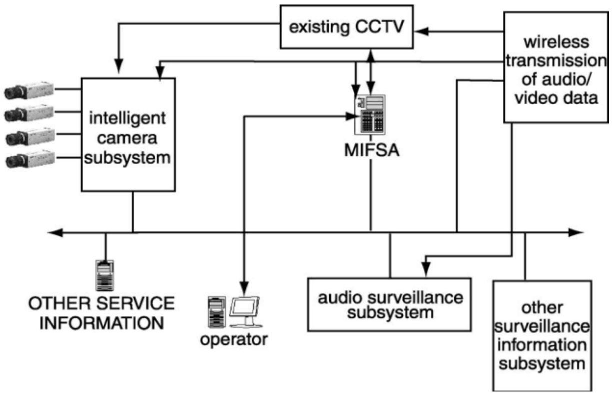

Fig. 1. An overview of PRISMATICA system [24]

Another impressive surveillance system is DARPA ARGUS-IS (autonomous real-time ground ubiquitous surveillance imaging system) [41]. With 1.8 Gigapixels video system, ARGUS-IS is able to auto-track every moving object within a 40 square kilometers (size of small city) using single platform. Such commercially available ISS products: DETEC AS (www.detec.no) and DETER (detection of events for threat evaluation and recognition) [42]. Intelligent surveillance system may play a significant role in security and safety in public, as well as in private domain. However, it is highly challenging due to some practical issues, such as:

- Performance: such as the system accuracy
- Robustness: the system should be robust again real word issues such as illumination variation, clutter, occlusion, weather change, camouflage, etc.
- Reliability
- Real time constrain: the system should fast enough
- Cost effective

3. Possible Sensor Modalities and Fusion Methods

3.1. Visible Camera

Visible (Video) camera is common sensor modalities for surveillance system. It has long been in use to monitor environments, people, events and activities. It is the most commercially available surveillance sensors starting from low cost IP camera to high performance professional CCTV. Security cameras have been placed in everywhere, from private homes, streets, public buildings, as well as in border between countries. Extensive research has been conducted for visible or video surveillance system [43]. Different types of visible cameras have been investigated for surveillance system such as color (or RGB) camera, monocular, stereo, omnidirectional camera, etc. Valera *et al* [44] divided the technological evolution of visual surveillance systems into three generations: Analog CCTV systems (1st generation), automated visual surveillance by combining computer vision technology with CCTV systems (2nd) and automated wide-area surveillance system (3rd).

3.2. Infrared (IR) and Thermal Camera

Visible camera is working well only in the environment that has enough illumination or light intensity, for example during in daytime. In the environment with low light intensity or during the night, visible camera cannot capture the scene effectively. In this case, there are two possible solutions: using infrared camera or thermal camera. Object (like human) that has contras temperature with the surrounding environment is much easier to distinguish in the thermal or infrared camera image than in the visible camera image.

Both cameras capture infrared radiation that is invisible for human eye, therefore the “infrared camera” and “thermal camera” terms are usually interchangeable. However, infrared camera usually referred to a camera that captures near-infrared (NIR) or short-wavelength infrared (SWIR) emissions to increase the visibility. Infrared cameras are suitable for environments with a low illumination level. While thermal camera is referred to a camera that is able to capture long-wave or far-infrared (FIR) radiation emitted or reflected by objects. Thermal camera is useful if the scene is completely dark. Thermal camera can be divided into two types: cooled and uncooled. Cooled thermal camera provides higher resolution and image quality, but generally more expensive and consumes more power. Examples of infrared and thermal camera are FLIR cameras, produced by FLIR System (www.flir.com), and AXIS Q19 camera series. Fig. 2 shows a scene that is captured using visible camera and thermal camera in the same time [18].


Fig. 2. A scene captured using visible and thermal camera [18]

3.3. Radar and lidar

Range sensing is an interesting sensor modality due to its accuracy, large field of view and robustness with respect to illumination changes. Such range sensing includes radar (radio detection and ranging) and lidar (light detection and ranging). Radar uses uses radio waves for sensing, while lidar uses light or laser. In range data, changes in the background can be easily filtered out by excluding all data outside of the tracking area. One drawback is that range data is generally less informative than vision data for person or object recognition. Spinello *et al* [21-22] proposed people tracking using 3D lidar. Recenltly, Banedek [23] also proposed 3D people surveillance using rotating multi-beam (RBM) lidar (as shown in Fig. 3). Javed *et al* [19] develop automatic target classifier (such pedestrian and vehicles) using ground surveillance radar. While Kocur *et al* in [20] used ultra-wideband (UWB) radars for surveillance robot.

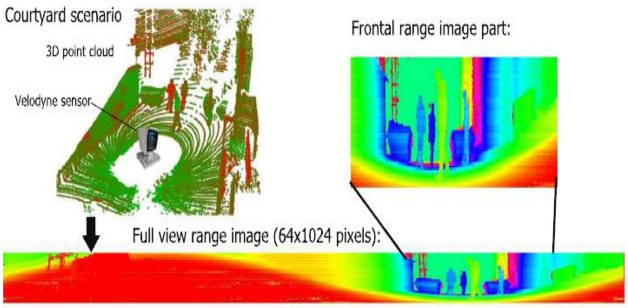

Fig. 3. People surveillance using 3D lidar [23]

3.4. Other sensor (Audio, Ultrasonic, etc)

There are many sensor modalities have been explored to improve or assist surveillance system, such as: audio 24], ultrasonic [30], passive infrared (PIR) [31] pressure sensor, etc. Environmental sound like breaking of glass, dog's barking, people screaming, fire alarm, gun firing and similar kind of sounds, may give a reasonable degree of confidence in making a decision about 'secure' or 'insecure' state [45]. Some sensors maybe used for alerting; ones they detect an object, visible camera (or other sensor) is activated for more reliable recognition. Bai *et al* [30] developed an embedded home surveillance system based on multiple ultrasonic sensors. In their other work [31], they used pyroelectric infrared sensors (PIR) and pressure sensors as an alert system to save the power.

3.5. Sensor Fusion

Intuitively, combining multiple sensors will provide more accurate information about the targeted object. Multiple sensors might be homogeneous (same modality, such as multiple cameras) or heterogeneous (different modalities). Some sensor modalities are intuitively closed and complementary, such as visible, infrared and thermal camera, since they capture information in 2D image perspective. Similarly, both lidar and radar capture information in range domain (2D or 3D). Sensor fusion may happen at low-level (data fusion), high level (decision fusion) or in between. In data fusion, each sensor sends its original measurement to the fusion center, and then the center makes the decision about the event. In decision fusion, each sensor makes its own decision based on its own measurement, and then the fusion center makes the final decision based on all individual decisions (for example using majority voting). Each fusion scenario has its own advantages and drawbacks. Challenges in sensor fusion: how to handle different data modalities (visual, audio, radio signal, etc), data imperfection, conflicting data, sensor topology etc [28]. More information about basic sensor fusion may refer to [25-29].

Extensive works have been done on visual surveillance system using multiple cameras [46-48]. Reference [29] reviews recent progress in intelligent video surveillance using multiple cameras that include multi-camera calibration, computing the topology of camera networks, multi-camera tracking, object re-identification, and also multi-camera activity analysis. Robertson *et al* [49] combine visible, infrared and thermal camera for outdoor people detection for moving platform (vehicle). Premebida *et al* [50] proposed pedestrian detection combining RGB camera and dense LIDAR data

4. Data Processing Techniques for ISS

4.1. Foreground-Background Segmentation

Foreground-background segmentation is the first important step for intelligent surveillance system. The goal is to separate the object or moving object (foreground) and the environment (background). It commonly referred also as background modeling, background subtraction, or change detection. Many foreground-background segmentation techniques have been proposed, especially for visible/video surveillance. Several review paper also available that focused discuss foreground-background segmentation methods [2-7]. Bouwmans [5] discussed and provide a comprehensive list most of available techniques (see Table 1). Fig. 4 shows foreground-background segmentation using different methods: generalized mixture of Gaussians (MOG) [51], non parametric kernel density estimation [52], and codebook [53].

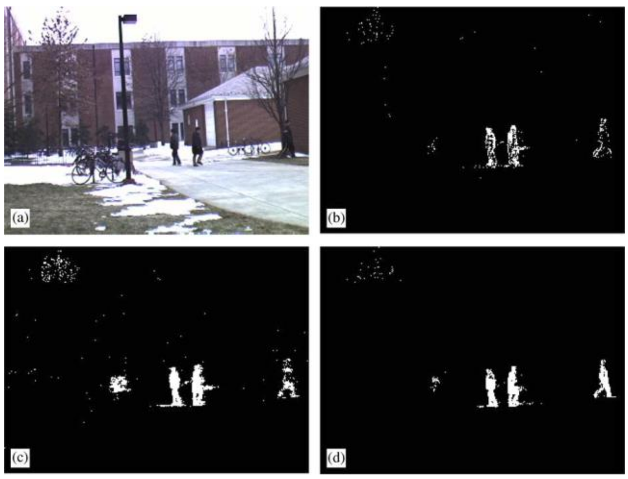

Fig. 4. Foreground-background segmentation: (a) original image, (b) MOG, (c) Kernel, (d) Codebook [53]

Table 1. Background Modeling Methods (reproduced from [Bouwmans2011])

Category	Methods	Main Contributor (Author, year)
Basic Background Modeling	Mean Median Histogram over time	Lee *et al.* (2002) [54] Mac Farlane *et al.* (1995) [55] Zheng *et al.* (2006) [56]
Statistical Background Modeling	Single Gaussian Mixture of Gaussians Kernel Density Estimation	Wren *et al.* (1997) [57] Stauffer and Grimson (1999) [51] Elgammal *et al.* (2000,2002) [52,58]
Fuzzy Background Modeling	Fuzzy Running Average Type-2 Fuzzy Mixture of Gaussians	Sigari *et al.* (2008) [59] El Baf *et al.* (2008) [60]
Background Clustering	K-Means Codebook	Butler *et al.* (2003) [61] Kim *et al.* (2005) [53]
Neural Network Background Modeling	General Regression Neural Network Self Organizing Neural Network	Culibrk *et al.* (2006) [62] Maddalena and Petrosino (2007) [63]
Wavelet Background Modeling	Discrete Wavelet Transform	Biswas *et al.* [64]
Background Estimation	Wiener Filter Kalman Filter Tchebychev Filter	Toyama *et al.* (1999) [65] Messelodi *et al.* (2005) [66] Chang *et al.* (2004) [67]

Obviously any possible techniques for foreground-background segmentation depend on the corresponding sensor modality. Recently, Sobral *et al* [7] compared 29 methods using BMC (Background Models Challenge) dataset [68]. Top five promising methods based on this experimental work are the methods that proposed by Wren *et al* [57], Kaewtrakulpong *et al* [69], Yao *et al* [70], Maddalena *et al* [63], Hofmann *et al* [71]. Cristani *et al* [3] discussed also other sensing modalities (such as audio, infrared and thermal camera) in their survey paper. Most of the proposed method in background-foreground segmentations employed only single sensor modality, and particularly using visible camera. Obviously, combining different sensor modality would make the system more robust or simplify the processing process for segmentation. For example, by combining visible camera and range data the background-foreground segmentation task become easier. Changes in the background can be easily filtered out by excluding all data outside of the observed area in the range data and the visible/image data is used for fine segmentation.

4.2. Object Detection and Classification

The ability to automatically detect and classify object (such as human and vehicle) is one of key component in intelligent surveillance system (ISS). For a machine (computer), detecting object like human is a hard job due to wide range of possible appearance as result of changing articulated pose, clothing, lighting and background [8]. Huge methods have been proposed for people detection based on visual camera. In their experimental survey, Enzweiler *et al* [9] showed an advantage of HOG/linSVM [72] at higher image resolutions and lower processing speeds, and a superiority of the wavelet-based AdaBoost [73] cascade approach at lower image resolutions and closed real-time processing speeds. In the more recent benchmarking effort, Dollar *et al* [10] show that FPDW [74] has the best overall performance, but If computational cost is not a consideration, then MULTIFTR+MOTION [75] is the best choice.

Spinello *et al* [21, 22] proposed people detection using a bottom-up top-down detector, based on lidar data. The bottom-up detector learns a layered person model from a bank of specialized classifiers for different height levels of people that collectively vote into a continuous space. In the top-down step, the candidates are classified using features that are computed in voxels of a boosted volume tessellation. While in [23] Banedek *et al* map the 3D lidar point data into dept-image, and performing people detection in 2D. Spinello *et al* in [22] presented a people detection approach based on RGB-Depth sensors that provide both image and range data. Fig. 5 shows example of their result for people detection. Most of the proposed method in objects detection and classification focusing only for couple of object types for example human and car. In fact, in the real setting, there are a lot of object that should be considered also for example different type of animal or other subject that have potential threat for security or safety.

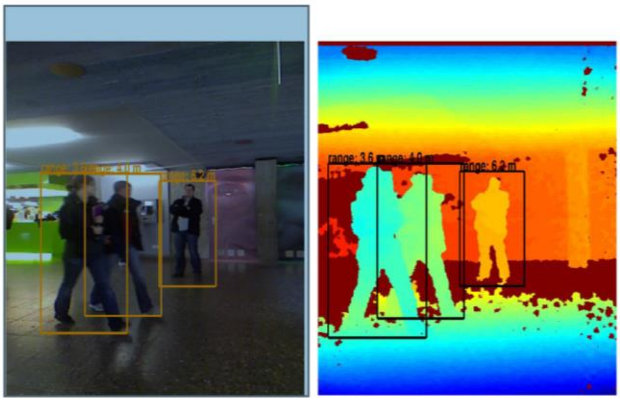

Fig. 5 people detection using RGB-D: color image (left), depth image (right)

4.3. Object Tracking and Re-Identification

After object detection, surveillance systems generally track the object in the spatiotemporal domain. Object tracking in are realistic scene is a challenging problem due to illumination changes, occlusion, clutter, sensor motion, and other issues. A large number of visual tracking algorithms (based on visible camera) have been proposed in recent years. Object tracking methods based on visual camera can be classified into five groups: model-based, appearance-based, contour- and mesh-based, feature-based, and hybrid methods [76]. Several review papers that focused on visual tracking problem are available such as [11-14]. Recently, Smeulders *et al* [12] performed experimental survey based on Amsterdam Library of Ordinary Videos (ALOV) for 19 online trackers. Another effort for benchmarking visual object trackers was proposed by Wu *et al* [14]. According to the Visual Object Tracking challenge (VOT2014) result, the best tracker (combined accuracy and robustness) is the discriminative scale space tracker (DSST) proposed by [77]. This tracker extended the minimum output sum of squared errors (MOSSE) tracker [78] with robust scale estimation.

Recently, some attempts have been done for people tracking using other than visible camera, such as using radar, lidar etc. For example, Mitzel *et al* [79] using stereo range data for real-time multi-person tracking. They did not only analyze 2D image, but also the range information from stereo camera. Fig. 6 shows an example of their results. Javed *et al* [19] develop automatic target classifier (such pedestrian and vehicles) using ground surveillance radar. While Kocur *et al* in [20] used ultra-wideband (UWB) radars for surveillance robot. Based on lidar data, Spinello *et al* [21, 22] proposed 3D people tracking using multi-target multi-hypothesis tracking approach. For people detection they employed a bottom-up top-down detector (explained in the previous section). Banedek *et al* [23] proposed an approach on real-time 3D people surveillance, with probabilistic foreground modeling, multiple persons tracking and on-line re-identification. The tracker module was also tested in real outdoor scenarios, with multiple occlusions and several re-appearing people during the observation period.

Fig. 6. Automatic people tracking [79]

4.4. Behavioral Analysis

There is an increasing interest to automatically analyze surveillance scene not only in the "object level" (such as detecting, tracking), but also further into "event level". Particular interests such are automated human behavior analysis [80, 81], group behavior analysis [82], crown analysis [36, 83, 84], and event analysis. Some review papers have been devoted to this topic [15-17]. Human behavior analysis can play a significant role in security by decreasing the time taken to thwart unwanted events and picking them up during the suspicion stage itself [17]. Analysis of human behavior although crucial, is highly challenging. Basic component in human behavior analysis is classifying the human behavior. Different ways have been proposed to classify human behavior. Kiryati *et al* [85] proposed simple classification: normal and abnormal. Foroughi *et al* [86] expand the classification into normal, unusual and abnormal. Previously, Park and Aggarwal [87] classified the activities as positive, neutral and negative activities.

Human can be monitored as isolated individuals, groups of people, or crowds. Examples of group events are people fighting, people being followed, people walking together, terrorists launching attacks in groups, etc [82]. Solmaz *et al* [83] proposed a method for identifying five crowd behaviors (bottlenecks, fountainheads, lanes, arches, and blocking) in visual scenes. Bremond *et al* [88] proposed an activity-monitoring framework for recognizing behaviors, involving either isolated individuals, groups of people, or crowds, in the context of visual monitoring of metro scenes, using multiple cameras. For example, Fig. 7 shows their result to recognize the "fighting behavior" in a metro station. The combined four methods or descriptions to recognize fighting behavior such as: (A) a group of people gathering around a lying person, (B) group width varied significantly, (C) people inside a group separate quickly, and (D) group trajectory changes very fast [88].

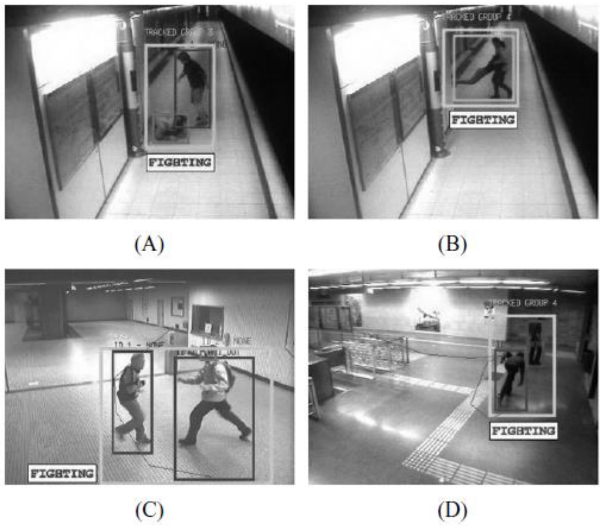

Fig. 7. Recognizing "fighting behavior" in a metro station [88]

However, it should be noted that current research in the behavior analysis are still considering simple or simplified scene. More realistic and complex scene should be investigated. For example, in real "fighting behavior" it may be involving also the use of weapon such as knife or gun, so there will be less contact between fighting groups or persons. The four descriptions proposed by Bremond *et al* above may fails to characterize this fighting behavior.

5. Conclusion and Future Direction

In this paper, general overview of intelligent surveillance systems has been presented. Such intelligent systems are promising to be implemented in various environments and applications. This paper also has discussed some possible sensor modalities and their fusion scenarios to improve the system performance. Numerous techniques have been proposed to tackle several main processing steps: background-foreground segmentation, object detection and classification, tracking, and behavioral analysis. Although several promising results have been obtained, further studies are needed for real implementation with more complex settings. For example, in the background-foreground segmentation process, different combination of sensor modality should be explored to make the system robust or to simplify the processing process. Current studies in behavior analysis are still considering simplified scene, and thus more realistic and complex scene should be investigated. With decreasing price in sensor and processing devices, researchers should also consider investigating and developing a low cost intelligent surveillance system.

References

1. N. Sulman, T. Sanocki, D. Goldgof and R. Kasturi, *How effective is human video surveillance performance?*, 19th International Conference on Pattern Recognition, Tampa, FL, 2008, pp.1-3R.
2. R.J. Radke, S. Andra, O. Al-Kofahi and B. Roysam, *Image change detection algorithms: A systematic survey*, IEEE Trans. Image Process. 14 (2005) 294–307.
3. M. Cristani, M. Farenzena, D. Bloisi and V. Murino, *Background Subtraction for Automated Multisensor Surveillance: A Comprehensive Review*, EURASIP Journal on Advances in Signal Processing 2010 (2010) 343057.
4. T. Bouwmans, F. El Baf, and B. Vachon, *Statistical background modeling for foreground detection: A survey*, Handbook of Pattern Recognition and Computer Vision (Volume 4). Singapore: World Scientific, 2010.
5. T. Bouwmans, *Recent advanced statistical background modeling for foreground detection: A systematic survey*, Recent Patents Comput. Sci. 4 (2011) 147–176.
6. S. Brutzer, B. Hoferlin, G. Heidemann, *Evaluation of background subtraction techniques for video surveillance*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, USA, 2011, pp. 1937-1944.
7. A. Sobral, A.Vacavant, *A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos*, Computer Vision and Image Understanding 122 (2014) 4-21.
8. M. Paul, S. M. Haque and S. Chakraborty, *Human detection in surveillance videos and its applications-a review*, EURASIP Journal onar Advances in Signal Processing 2013(2013) 1-16.
9. M. Enzweiler and D. M. Gavrila, *Monocular pedestrian detection: Survey and experiments*, IEEE Trans. Pattern Analysis and Machine Intelligence 31 (2009) 2179-2195.
10. P. Dollar, C. Wojek, B. Schiele and P. Perona, *Pedestrian Detection: An Evaluation of the State of the Art*, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (2012) 743-761.
11. A. Yilmaz and M. Shah, *Object tracking: A survey*, Journal ACM Computing Surveys 38 (2006) article 13 (45 pages).
12. A.W.M. Smeulders, Dung M. Chu, R.Cucchiara, S. Calderara, A. Dehghan and M. Shah, *Visual Tracking: an Experimental Survey*, IEEE Transaction on Pattern Analysis and Machine Intelligence 36 (2014) 1442-1468.
13. A. Bedagkar-Gala and Shishir K. Shah, *A survey of approaches and trends in person re-identification*, Image and Vision Computing 32 (2014) 270-286.
14. Y. Wu, J Lim and M Yang, *Object Tracking Benchmark*, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (2015) 1834-1848.
15. T. Ko, *A survey on behavior analysis in video surveillance for homeland security applications*, 37th IEEE Applied Imagery Pattern Recognition Workshop, Washington, DC, USA, 2008, pp.1-8.
16. M. Cristani, R. Raghavendra, A. Del Bue and V. Murino, *Human behavior analysis in video surveillance: A Social Signal Processing perspective*, Neurocomputing 100 (2013) 86-97.
17. D. Gowshikaa, S Abirami and R Baskaran, *Automated Human Behaviour Analysis from Surveillance videos: A survey*, Artificial Intelligence Review 42 (2014) 747-765.
18. R. Gade and T. B. Moeslund, *Thermal cameras and applications: A survey*, Machine Vision & Applications 25 (2014) 245-262.
19. A. Javed, A. Ejaz, S. Liaqat, A. Ashraf and M.B. Ihsan, *Automatic target classifier for a Ground Surveillance Radar using linear discriminant analysis and Logistic regression*, Radar Conference (EuRAD), 2012 9th European, Amsterdam, 2012, pp.302-305.
20. D. Kocur, P. Kazimir, J. Fortes, D. Novak, M. Drutarovsky, P. Galajda and R. Zetik, *Short-range UWB radar: Surveillance robot equipment of the future*, IEEE International Conference on Systems, Man and Cybernetics (SMC), San Diego, CA, USA, 2014, pp.3767-3772.
21. L. Spinello, K. O. Arras, R. Triebel and R. Siegwart, *A layered approach to people detection in 3D range data*, AAAI Conference on Artificial Intelligence (AAAI-10), Atlanta, Georgia, USA, 2010.
22. L. Spinello, M. Luber and K.O. Arras, *Tracking people in 3D using a bottom-up top-down detector*, 2011 IEEE International Conference on Robotics and Automation (ICRA), Shanghai, 2011, pp.1304-1310.
23. Csaba Benedek, *3D people surveillance on range data sequences of a rotating Lidar*, Pattern Recognition Letters 50 (2014) 149-158.
24. Benny Ping Lai Lo, Jie Sun and Sergio A.Velastin, *Fusing Visual and Audio Information in a Distributed Intelligent Surveillance System for Public Transport Systems*, Acta Automatica Sinica, 29 (2003) 393-407.
25. D.L. Hall and J. Llinas, *An introduction to multisensor data fusion*, Proceedings of the IEEE 85 (1997) 6-23.
26. E. F. Nakamura, A. A. F. Laureiro and A. C. Frery, *Information fusion for wireless sensor networks: Methods, models, and classifications*, ACM Comput. Surveys 39 (2007) 1–55.
27. E.I. Gokce, A.K. Shrivastava, Jung Jin Cho and Yu Ding, *Decision Fusion from Heterogeneous Sensors in Surveillance Sensor Systems*, IEEE Transactions on Automation Science and Engineering 8 (2011) 228-233.
28. B. Khaleghi, Alaa Khamis, Fakhreddine O. Karray and Saiedeh N. Razavi, *Multisensor data fusion: A review of the state-of-the-art*, Information Fusion 14 (2013) 28-44.
29. X. Wang, *Intelligent multi-camera video surveillance: A review*, Pattern Recognition Letters, 34 (2013) 3-19.
30. Y.W. Bai, Li-Sih Shen and Zong-Han Li, *Design and implementation of an embedded home surveillance system by use of multiple ultrasonic sensors*, IEEE Transactions on Consumer Electronics, 56 (2010) 119-124.
31. Y.W. Bai, Z. l. Xie and Z. H. Li, *Design and implementation of a home embedded surveillance system with ultra-low alert power*, IEEE Transactions on Consumer Electronics, 57 (2011) 153-159.
32. J.L. Castro, M. Delgado, J. Medina and M.D. Ruiz-Lozano, *Intelligent surveillance system with integration of heterogeneous information for intrusion detection*, Expert Systems with Applications, 38 (2011) 11182-11192.
33. A. Prati, R. Cucchiara and R. Vezzani, *A multi-camera vision system for fall detection and alarm generation*, Expert Systems, 24(5) 334–345.
34. F. Porikli, Y. Ivanov and T. Haga, *Robust abandoned object detection*

using dual foregrounds, EURASIP Journal on Advances in Signal Processing, (2008) article no. 30.

35. Y.L. Chen, Bing-Fei Wu, Hao-Yu Huang and Chung-Jui Fan, *A Real-Time Vision System for Nighttime Vehicle Detection and Traffic Surveillance*, IEEE Transactions on Industrial Electronics 58 (2014) 2030-2044.
36. J.C.S. J. Junior, S. R. Musse and C. R. Jung, *Crowd analysis using computer vision techniques: A survey*, IEEE Signal Processing Magazine 27 (2010) 66-77.
37. R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto and O. Hasegawa, *A system for video surveillance and monitoring*, Tech. Rep. CMU-RI-TR- 00-12, Robotics Institute, Carnegie Mellon University (2000).
38. I. Haritaoglu, D. Harwood and L.S. Davis, *W4: Real-time surveillance of people and their activities*, Patter Analysis and Machine Intelligence, 22 (2000) 809–830.
39. S.A. Velastin, L. Khoudour, B.P.L Lo, J. Sun and M.A. Vicencio-Silva, *PRISMATICA: a multi-sensor surveillance system for public transport networks*, 12th IEE International Conference on Road Transport Information and Control, London, UK, 2004, pp.19-25.
40. N. T. Siebel and S. Maybank, *The ADVISOR visual surveillance system*, Proc. ECCV 2004 Workshop Applications of Computer Vision, Prague, Czech Republic, 2004, pp.103-111.
41. B. Leininger, *et al. Autonomous real-time ground ubiquitous surveillance-imaging system (ARGUS-IS)*. Proc. SPIE 6981 (2008) 69810H.
42. V. Morellas, I. Pavlidis and P. Tsiamyrtzis, *DETER: Detection of Events for Threat Evaluation and Recognition*, Machine Vision and Applications, 15(2003) 29-46.
43. J. S. Kim, D. H. Yeom and Y. H. Joo, *Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems*, IEEE Transactions on Consumer Electronics 57 (2011) 1165-1170.
44. M. Valera and S. A. Velastin, *Intelligent distributed surveillance systems: a review*, IEEE Proceedings - Vision, Image and Signal Processing, vol. 152 (2005) 192-204.
45. P. Kumar, A. Mittal and P. Kumar, *A Multimodal Audio, Visible and Infrared Surveillance System (MAVISS)*, 3rd IEEE International Conference on Intelligent Sensing and Information Processing (ICISIP), Bangalore, Karnataka, India, 2005, pp. 151-157.
46. Q. Cai and J.K. Aggarwal, *Tracking human motion in structured environments using a distributed-camera system*, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (1999) 1241-1247.
47. R.T. Collins, A.J. Lipton, H. Fujiyoshi and T. Kanade, *Algorithms for cooperative multisensor surveillance*, Proceedings of the IEEE 89 (2001) 1456-1477.
48. X. Wang, S. Wang and D. Bi, *Distributed Visual-Target-Surveillance System in Wireless Sensor Networks*, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 39 (2009) 1134-1146.
49. N.M. Robertson and J. Letham, *Contextual person detection in multi-modal outdoor surveillance*, Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European, Bucharest, Romania, 2012, pp.1930-1934.
50. C. Premebida, J. Carreira, J. Batista and U. Nunes, *Pedestrian detection combining rgb and dense lidar data*, International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 2014, pp. 4112-4117.
51. C. Stauffer and W. Grimson, *Adaptive background mixture models for real-time tracking*, CVPR 1999 (1999) 246-252.
52. A. Elgammal, D. Harwood and L.Davis, *Non-parametric Model for Background Subtraction*, ECCV, Dublin, Ireland, 2000, pp. 751-767.
53. K. Kim, T. Chalidabhongse, D. Harwood and L. Davis, *Real-time Foreground-Background Segmentation using Codebook Model*, Real-Time Imaging 11 (2005) 172-185.
54. B. Lee and M. Hedley, *Background Estimation for Video Surveillance*, Image and Vision Computing New Zealand (IVCNZ 2002), New Zealand, 2002, pp. 315-320.
55. N. McFarlane and C. Schofield, *Segmentation and tracking of piglets in images*, BMVA 1995 (1995) 187-193.
56. J. Zheng, Y. Wang, N. Nihan and E. Hallenbeck, *Extracting Roadway Background Image: A mode based approach*, Journal of Transportation Research Report 1944 (2006) 82-88.
57. C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, *Pfinder: Real-Time Tracking of the Human Body*, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 780-785.
58. A. Elgammal, R. Duraiswami, D. Harwood and L. S. Davis, *Background and foreground modeling using nonparametric kernel density estimation for visual surveillance*, Proceedings of the IEEE 90 (2002) 1151-1163.
59. M. Sigari,N. Mozayani and H. Pourreza, *Fuzzy Running Average and Fuzzy Background Subtraction: Concepts and Application*, International Journal of Computer Science and Network Security 8 (2008) 138-143.
60. F. El Baf, T. Bouwmans and B. Vachon, *Type-2 fuzzy mixture of Gaussians model: Application to background modeling*, ISVC 2008, Las Vegas, USA, 2008, pp. 772-781.
61. D. Butler and S. Sridharan, *Real-Time Adaptive Background Segmentation*, International Conference on Multimedia and Expo, Baltimore, Maryland, 2003, pp. III-341-4.
62. D. Culbrik, O. Marques, D. Socek, H. Kalva and B. Furht, *Neural network approach to background modeling for video object segmentation*, IEEE Transaction on Neural Networks 18 (2007) 1614–1627.
63. L. Maddalena and A. Petrosino, *A self-organizing approach to background subtraction for visual surveillance applications*, IEEE Transactions on Image Processing 17 (2008) 1729–1736.
64. S. Biswas, J. Sil and N. Sengupta, *Background Modeling and Implementation using Discrete Wavelet Transform: a Review*, JICGST - GVIP 11 (2011) 29-42.
65. K. Toyama, J. Krumm, B. Brumitt and B. Meyers, *Wallflower: Principles and Practice of Background Maintenance*, International Conference on Computer Vision, Corfu, Greece, 1999, pp. 255-261.
66. S. Messelodi, C. Modena, N. Segata and M. Zanin, *A Kalman filter based background updating algorithm robust to sharp illumination changes*, ICIAP 2005, Cagliari, Italy, 2005, pp. 163-170.
67. R. Chang, T. Ghandi and M. Trivedi, *Vision modules for a multisensory bridge monitoring approach*, The 7th International IEEE Conference on, Intelligent Transportation Systems, Washington, WA, USA, 2004, pp. 971-976.
68. A. Vacavant, T. Chateau, A. Wilhelm and L. Lequièvre, *A benchmark dataset for foreground/background extraction*, Background Models Challenge (BMC) at Asian Conference on Computer Vision (ACCV), Daejeon, Korea, 2012, pp. 291–300.
69. P. Kaewtrakulpong and R. Bowden, *An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection*, Video Based Surveillance Systems: Computer Vision and Distributed Processing, US: Springer US, 2001.
70. J. Yao and J. marc Odobez, *Multi-layer background subtraction based on color and texture*, IEEE Computer Vision and Pattern Recognition Conference (CVPR), Minneapolis, MN, USA, 2007, pp. 1 - 8.
71. M. Hofmann, P. Tiefenbacher and G. Rigoll, *Background segmentation with feedback: The pixel-based adaptive segmenter*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition

Workshops (CVPRW), Providence, RI, USA, 2012, pp. 38–43.

72. N. Dalal and B. Triggs, *Histograms of Oriented Gradients for Human Detection*, IEEE Int'l Conf. Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, pp. 886-893.
73. P. Viola, M. Jones and D. Snow, *Detecting Pedestrians Using Patterns of Motion and Appearance*, Int'l J. Computer Vision, 63 (2005) 153-161.
74. P. Dollar, S. Belongie and P. Perona, *The Fastest Pedestrian Detector in the West*, Proc. British Machine Vision Conf., 2010, pp. 68.1-68.11.
75. S. Walk, N. Majer, K. Schindler and B. Schiele, *New Features and Insights for Pedestrian Detection*, IEEE Conf. Computer Vision and Pattern Recognition, San Francisco, CA , USA, 2010, 1030 - 1037.
76. A. Cavallaro, O. Steiger and T. Ebrahimi, *Tracking video objects in cluttered background*, IEEE Transactions on Circuits and Systems for Video Technology 15 (2005) 575-584.
77. M. Danelljan, G. Hager, F.S. Khan and M. Felsberg, *Accurate scale estimation for robust visual tracking*, Proceedings of the British Machine Vision Conference BMVC, Nottingham, UK, 2014, pp - .
78. D.S. Bolme, J.R. Beveridge, B.A. Draper and Y.M. Lui, *Visual object tracking using adaptive correlation filters*, Comp. Vis. Patt. Recognition (2010), San Francisco, CA, USA, 2010, pp. 2544-2550.
79. D. Mitzel and B. Leibe, *Real-time multi-person tracking with detector assisted structure propagation*, Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, Barcelona, 2011, pp. 974-981.
80. W. Li, Z. Zhang and Z. Liu, *Action recognition based on a bag of 3D points*, IEEE computer society conference on computer vision and pattern recognition workshops, San Francisco, CA, USA, 2010, pp. 9–14.
81. C. Rougier, J. Meunier, A. St-Arnaud and J. Rousseau, *Robust video surveillance for fall detection based on human shape deformation*, IEEE Trans Circuits Syst Video Technol 21(2011) 611–622.
82. W. Lin, M.T. Sun, R. Poovendran and Z. Zhang, Group event detection with a varying number of group members for video surveillance. IEEE Trans Circuits Syst Video Technol 20(8):1057–1067, 2010.
83. B. Solmaz Brian E. Moore and Mubarak Shah, *Identifying Behaviours in Crowd Scenes Using Stability Analysis for Dynamical Systems*, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (2012) 2064-2070.
84. H. Idrees, Nolan Warner and Mubarak Shah, *Tracking in dense crowds using prominence and neighborhood motion concurrence*, Image and Vision Computing 32 (2014) 14-26.
85. N. Kiryati, T.R. Riklin, Y. Ivanchenko and S. Rochel, *Real-time abnormal motion detection in surveillance video*, IEEE 19th International Conference on Pattern Recognition, Tampa, Florida, USA, 2008, pp. 1–4.
86. H. Foroughi, H.S. Yazdi, H. Pourreza and M. Javidi, *An eigenspace-based approach for human fall detection using integrated time motion image and multi-class support vector machine*, IEEE 4th international conference on intelligent computer communication and processing, Cluj-Napoca, 2008, pp. 83–90.
87. S. Park and J.K. Aggarwal, *Semantic-level understanding of human actions and interactions using event hierarchy*, IEEE computer society conference on computer vision and pattern recognition workshops, Washington, DC, USA ,2004, pp 12.
88. F. Bremond, M. Thonnat and M. Zuniga, *Video understanding framework for automatic behavior recognition*, Behav Res Methods 38(2005) 416–426.