A Swift and Novel Backward Modeling Based Video Tracking Using Image Fusion

Ritu dixit[1] Mr. Rohit Miri[2]
[1]M.Tech Student [2]Assistant Professor
[1,2]Department of Computer science Engineering
[1,2]Dr. C.V. Raman University, Bilaspur, Chhattisgarh-India

Abstract— *We present a image fusion techniques to automatically video surveillance combine images of Human movement tracking. Automatic video surveillance of dynamic and complex scenes is currently one of the most active research topics in image fusion .The methods can also be used for practical applications. It aims to automatically detect, recognize and track people and objects from image sequences in order to understand and describe dynamics and interactions among them. While traditional passive video surveillance is ineffective when the number of cameras exceeds the ability of human operators to keep track of the evolving scene, multicamera automatic systems allowing for accurate event detection. The multi sensor camera is used for tracking. The matlab simulation is used to understand the video surveillance tracking of Human movement.*

Keywords— Video Tracking, Image Fusion, multicamera

I. Introduction

Automatic video surveillance of dynamic and complex scenes is one of the most active research topics in Computer Vision. It aims to automatically detect, recognize and track people and objects from image sequences in order to understand and describes dynamics and interactions among them .Computer vision and video based surveillance have the potential to assist in maintaining public safety and security. Virtually all public spaces and critical infrastructures in the European Community have a multiple sensor surveillance system installed, many of which claim to have automatic surveillance features.Typical application domains for video surveillance include public areas (city streets, school campuses, museums), transport (airports, train stations, underground, motorways), retail (theft prevention, understanding shopper behaviour), and financial institutions (banks and casinos)

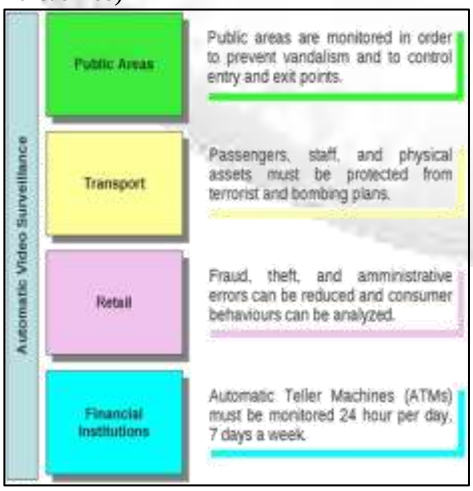

Fig. 1.1: Examples of business and security issues that can be managed by automatic video surveillane

Video surveillance can be used also for allowing elderly people to remain active and productive for longer, to continue and engage in the society with more accessible on-line services, and to enjoy a healthier and higher quality of life for longer. Examples of applications about safety and security are accident monitoring, visitor validation and activity detection.

A. Video surveillance systems:

Automatic video surveillance task can be broken down into a series of sub-problems:

1) Object detection and categorization which detects and classifies the interesting objects in the field of view of the camera.
2) Multi-Target Tracking (MTT) where the objective is to estimate the trajectories of targets, keeping the identification of each target;
3) MTT across cameras with the goal of tracking the targets while observing them through multiple overlapping or non-overlapping cameraswhich in turn feed into higher level scene analysis and/or behaviour analysis modules.

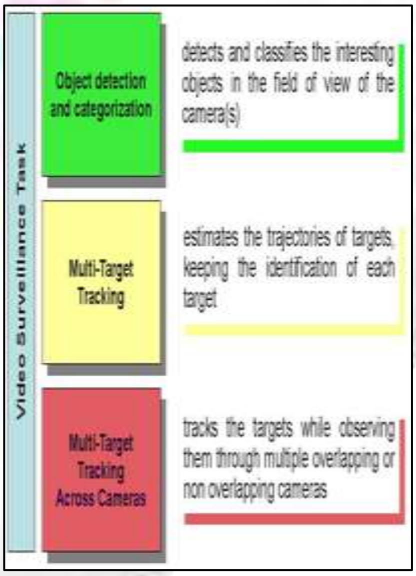

Fig. 1.2: Automatic video surveillance task

Furthermore, a video surveillance system should be able to react to particular events (e.g., we could want to send an alarm if an unauthorized vehicle enters a restricted area as depicted in Fig. 1.2)

B. General Scheme:

The general architecture of an automatic video surveillance system is depicted in Fig. 1.3. The majority of computer vision systems for surveillance are organized in a hierarchical way, with low level image processing techniques feeding into tracking algorithms which in turn feed into higher level scene analysis and/or behaviour analysis modules.\Object detection module aims to recognize moving objects in the scene (observations), while tracking module tries to determine correspondences between a set of observations separated over time. In a multi-camera environment, a data fusion module is required to manage data and to establish correspondences in a common coordinate framework in order to locate the same target in multiple views. Event understanding module classifies objects and analyses their dynamics in order to send an alarm if an event of interest is recognized.

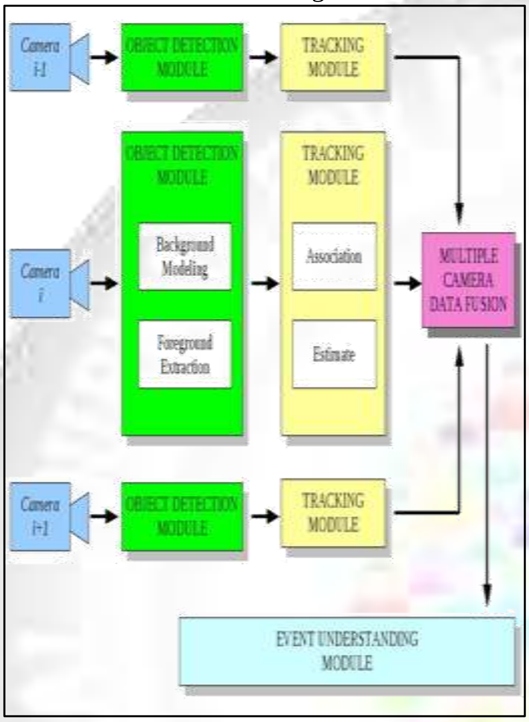

Fig. 1.3: A video surveillance system general scheme

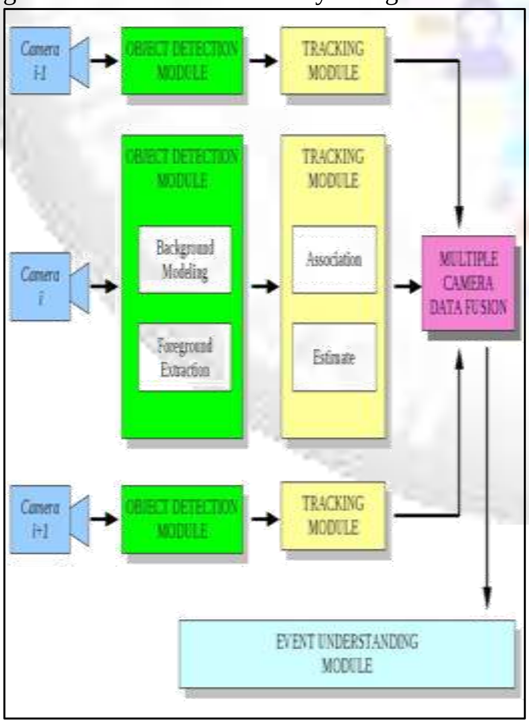

Fig. 1.4: An event to react to: Detect boats docking in the highlighted area.

C. Problem Identification:

Video surveillance can be used also for allowing elderly people to remain active and productive for longer, to continue and engage in the society with more accessible on-line services, and to enjoy a healthier and higher quality of life for longer. Examples of applications about safety and security are accident monitoring, visitor validation and activity detection.

II. LIMITATION EXISTING WORK

Traditional passive video surveillance has two main drawbacks:

1) Finding available human resources to observe the output is expensive;
2) Manual systems are ineffective when the number of cameras exceeds the ability of human operators to keep track of the evolving scene.

In large CCTV installations, with some hundreds of cameras, only a small fraction are monitored in real-time and the rest are only watched following an incident. The ratio of screens to cameras lies between 1:4 and 1:30, and the ratio of operatives to screens can be as high as 1:16.

Automatic systems can work 24 hours a day, 7 days a week allowing for accurate event detection and their cost is lower than maintaining a group of operatives.

However, none of the state-of-the-art system is able to deal with all the problems a monitored scene typically presents, namely occlusions, illumination changes, shadows, and crowded situations. Furthermore, due to sensor noise, motion in the scene and real-time constraints the objective of automatically detecting and tracking objects of interest is an hard task.

A winning strategy consists in integrating different techniques and data sources in an adaptive framework in order to exploit their advantages minimizing drawbacks. As an example, using a human operator as dynamic control agent that can provide updated contextual information about the observed environment can be exploited in improving system performance.

The majority of computer vision systems for surveillance purposes are made of four modules object detection, tracking, multi-camera Image fusion, and event understanding.

III. METHODOLOGY

Detection of object of interest in the scene is the stage in video surveillance systems. Regions corresponding to moving objects provide a focus of attention for later processes such as tracking and behaviour analysis because only these regions need be considered in the later processes. Background subtraction is a popular method for detecting moving objects from static cameras. A survey of state-of-the-art background subtraction algorithms is provided.

IV. BACKGROUND SUBTRACTION

Background subtraction (BS) is a widely used segmentation technique able to achieve real-time performance. BS aims to segment moving regions in image sequences comparing current frame to a model of the scene background. A pixel is classified as being from a moving object if the difference between the current frame and the background model is above a given threshold (see Fig. 4.1).

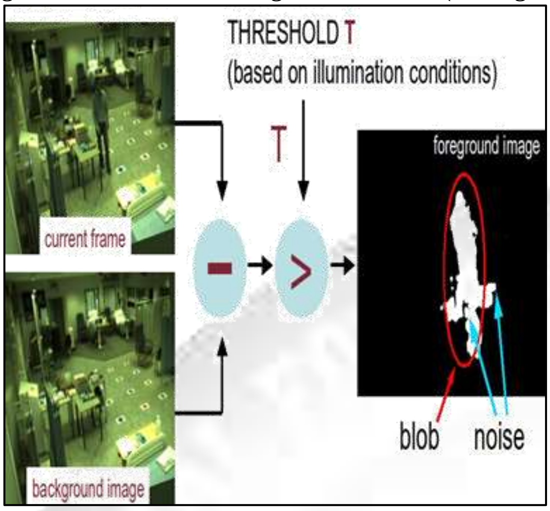

Fig. 4.1: Background subtraction method. Noise due to shadows is high-lighted.

How to create such a model is a challenging task due to illumination changes (gradual and sudden), shadows, camera jitter, and movement of background elements (e.g., trees swaying in the breeze, waves on water surfaces).

Background subtraction methods can be organized in: 1) per pixel, 2) per region, and 3) per frame. Per-pixel approaches is formed by methods that consider each pixel signal as an independent process. Region-based algorithms usually divide the frames into blocks and calculate block-speci c features in order to obtain the foreground. Frame-level class is formed by methods that look for global changes in the scene. Usually, they are used jointly with other pixel or region background subtraction approaches.

V. Tracking

Automatic surveillance systems generally track moving objects from one frame to another in an image sequence. The tracking process goal is to associate to the moving objects found by the segmentation process information like its identity, position, speed and acceleration. The ideal behaviour of a tracking system is to provide a set of tracks (possibly in an environment reference system) that are in a one-to-one correspondence with the objects appearing in the field of view of the sensor, i.e., no tracks associated with different objects and no multiple tracks associated with the same object. Factors that make this problem difficult include occlusion, complex light sources, and large size changes in the field of view.

Two major components can be distinguished in a typical visual tracker:

1) Target representation and localization is mostly a bottom-up process which has also to cope with the changes in the appearance of the target.
2) Filtering and data association is mostly a top-down process dealing with the dynamics of the tracked object, learning of scene priors, and evaluation.

The way the two components are combined and weighted is application dependent and plays a decisive role in the robustness and efficiency of the tracker.

The general Multi-Target Tracking (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.

VI. Image Fusion

A Image fusion system aims to address the problem of fusing information coming from different sensors in order to build a unified framework. This kind of process is also known as ltering and in general it implies using statistical information. Combining in a mathematically sound way such information requires to evaluate the uncertainty of measurements and then to incorporate the observations in the overall estimation.

VII. Multiple Camera Image Fusion

The need for using multiple cameras for monitoring and tracking arises for two reasons:

– Even if segmentation and tracking processes can be carried out by single-camera systems, wide area can be covered only through multiple-camera systems
– Multiple views can address the problem of handling occlusions and it is possible to retrieve 3D information about the detected objects.

In general, the use of multiple cameras brings a series of problems like camera installation, object matching, and data fusion. When dealing with multiple cameras, an important issue is the relationship between different views. It can be manually defined or computed automatically. The objective of multi-camera systems is to develop a model for establishing temporal and spatial correspondence of objects across different camera views

VIII. Camera Installation

Camera installation and positioning are not negligible issues in video surveillance systems development. Usually, in an indoor environment, it is more easy to choose the better position for each camera in order to minimize occlusions and blind spots. However, installing cameras in a controlled environment like a lab is di erent from installing them in a real-world environment like a supermarket or an airport, due to privacy, security and aesthetic issues.

Outdoor installation is more complex, due to the topography of the area (e.g., streets and tree lines) or city and building ordinances (e.g., historic building constraints).

IX. Camera Rectification

Camera calibration aims to specify the transformation between the camera and world reference frames. Orthogonal rectification can be used to cope with cam-era perspective. The map between a world plane and its perspective image is an homography. The world-plane to image-plane homography is fully defined by four points of which we know the relative position in the world plane. Once this homography is determined, the image can be back projected (warped) onto the world plane. This is equivalent to synthesize an image as taken from a front parallel view of the plane. This is known as orthogonal rectification of a perspective imageentities and their actions and interactions in the observed environment, to achieve much better situational awareness, and especially to recognize possible threats or dangerous conditions. Generally, these systems have to operate in crowded and complex environments where anomalous events have to be distinguished from the normal course of actions.

The complexity of the scenarios in which automatic surveillance systems of the next generation will have to operate requires an extension of the human role beyond passive monitor or partial controller to a dynamic control agent that can provide updated contextual information about the observed environment. The human user plays as the source that provides knowledge to the fusion system in an adaptive way. Information fusion between user data and visual data is performed in order to estimate and assess the situation at hand.

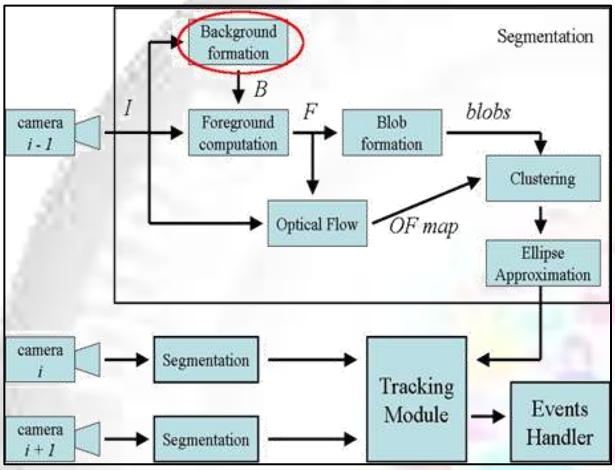

Fig. 4.1: A multi-camera video surveillance data own schema. Background modelling module is red circled.

X. Independent Multimodal Background Subtraction

A novel algorithm for creating a robust and multimodal back-ground model for accomplishing the background formation task is presented. Background modelling is the key step in background subtraction because BS is highly dependent on a good background model to reduce the in uence of changes in dynamic scenes derived from lighting and extraneous events as part of environment modelling.

The typical data ow for a multi-camera automatic video surveillance system is shown in Fig. 4.6 . Each camera provides a frame I that is used for computing the background model B and the foreground mask F. F is analysed in order to extract the blobs and to lter noise and false positives (e.g., exploiting the optical flow map). Usually the input for the tracking module is a set of bounding boxes (or ellipse) containing the blobs.

XI. Result & Discussion

Video surveillance systems have historically been evaluated in an unsystematic way, but carrying on a series of experiments is a key requirement in order to evaluate the accuracy of a system. It is possible to use:

1) Self-recorded data
2) Publicly available benchmark dataset.
3) Third party evaluation.

The best choice in order to evaluate an algorithm is to use benchmark data, which give the chance to quantitatively evaluate the performances of the method with respect to others. However, it is not always possible to use benchmarks (as an example, benchmarks for video surveillance using stereo cameras do not exist) and often benchmarks are recorded in controlled situations explicitly for a particular task (e.g., abandoned luggage detection). A third party evaluation can provide an objective evaluation of the system, but it is not always possible to perform, due to the requirement of developing a user interface to allow the test of event recognition performances.

A Image fusion scheme for improving performance through the use of a human operator as a dynamic control agent. The idea is to integrate hard electronically-based observational and soft human generated data in order to cope with the complexity of the scenarios in which automatic surveillance systems will have to operate. A set of key features for classifying automatic video surveillance systems has been provided. The final goal in automatic video Surveillance is to develop a system

1) able to model dynamic types of background,
2) able to track a great number of objects
3) that use multiple cameras for covering large areas and handling occlusions,
4) that can be valued using objective criteria.

The results shown in this thesis demonstrate that it is possible to develop a system near to achieve the features of an ideal system. Indeed, a wide area system running 24 hours per day in a challenging dynamical real-world scenario has been described. It is able to track hundreds of objects at the same time and an extensive third party evaluation using quantitative metrics demonstrates the effectiveness of the approach.

XII. Application and User Interface

In order to make available all the information gathered by the system in a useful way for the Human movement tracking, we have developed different visualizations of the results of the system. The main control window shows a animation of human movement tracking, integrating a map with the information about position and motion of the human. Figure : 5.1

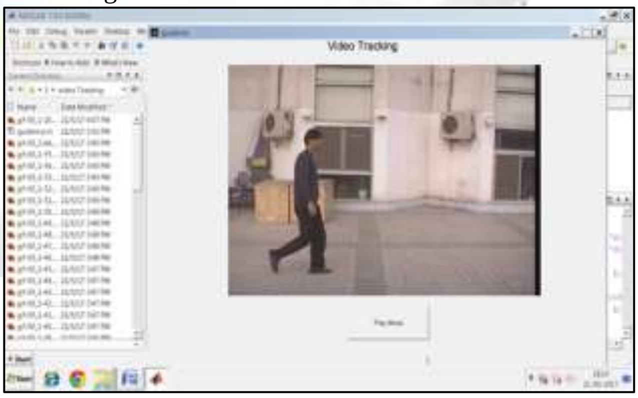

Fig. 5.1: Human Movement Recording using Multisensor Camera Image Fusion

Fig. 5.2: Human movement tracking using Independent multimodal Background Modelling

XIII. CONCLUSIONS

In this thesis automatic video surveillance systems have been analysed, giving an overview of the state-of-the-art as well as proposing novel approaches and solutions. Two real systems designed for different scenarios and applications have been described for Human movement monitoring in an outdoor scenario, for people tracking in an indoor crowded scenario monitored by stereo cameras. Moreover, an approach for dealing with conclusions in an outdoor environment with calibrated cameras has been analysed. Object detection, tracking, and Image fusion problems have been discussed and, for each of those topics, solutions applied to real systems have been described. A novel algorithms in the field of object detection has been detailed, Independent Multimodal Background Subtraction, a technique for computing a robust and accurate background gui model of the monitored scene.

XIV. FUTURE SCOPE

Significant progress has been achieved over the last decade in the field of automatic video surveillance, but successful cases are limited to controlled" situations, in which is possible to insert into the system strong knowledge about the environment to monitor (e.g., systems for highway monitoring).The main difficulty in realizing effective video surveillance systems that can reliably work in real conditions is the need of implementing techniques that are robust to the many different conditions that arise in real environments. Research challenges come with practical considerations such as the physical placement of cameras, the network bandwidth required to support them, installation costs, privacy concerns, and aesthetic constraints. Today, automatic video surveillance systems can detect pre-programmed events such as abandoned luggage, intrusion, and overcrowding. They can provide useful information about traffic statistics and people counting, but the challenge is in proving understanding at an higher level. As a short-term objective, the automatic video surveillance research community needs to agree on a set of predefined benchmark data to evaluate different approaches and choose the better ones among them. As a mid-term objective, a robust human computer interface may be developed in order to allow surveillance operatives helping the system in recovering from erroneous situations. Indeed, reasoning with uncertainty is likely to be the central feature for the next generation of automatic video surveillance systems. As a long-term objective, a formal and modular architecture of an ideal system must be specified in order to transform what are today ad hoc systems in standard compliance products.

REFERENCES

[1] Ahmad Taher Azar, Ali El Baz, Aboul Ella Hassanien,” Scientific Research Group in Egypt (SRGE) Faculty of computers and information, Benha University, Egypt

[2] Vaibhav R. Pandit ” Image Fusion in Remote Sensing Applications” Research Scholar S.G.B. Amravati University, Amravati. International Journal of Computer Applications (IJCA)(0975 – 8887) Volume 120 – No.10, June 2015

[3] electrical and computer science department of lehigh university ,Benmthlehem, pulished a book entitled as “Investigations of Image Fusion” in the year 2015. zhz3@lehigh.edu

[4] Shalima,“Review Of Image Fusion Techniques” Student, Department of Computer Science and Engineering, GNDU Amritsar, Punjab, India International Research Journal of Engineering and Technology (IRJET) ,ISSN: 2395 -0056 Volume: 02 Issue: 03 , June-2015.

[5] R.Johnson Suthakar, J.Monica Esther M.E, D.Annapoorani, F.Richard Singh Samuel ,” Study of Image Fusion- Techniques, Method and Applications” PG Student, Assistant Professor, PG Student, Department of Information Technology, Francis Xavier Engineering College, Tirunelveli, TamilNadu, India. Journal of Computer Science and Information Technology (IJCSMC), Vol. 3, Issue. 11, November 2014, pg.469 – 476 .

[6] Deron Rodrigues, Hasan Ali Virani, Shajahan Kutty “Multimodal Image Fusion Techniques for Medical Images using Wavelets”, Electronics and Telecommunication Department, Goa College of Engineering, Goa. International Journal of Research in Advent Technology, Vol.2, No.3, March 2014

[7] Deepak Kumar Sahu, M.P.Parsai,”Different Image Fusion Techniques –A Critical Review” ,Department of Electronics & Communication Engineering, Jabalpur Engineering College, Jabalpur MP, India International Journal of Modern Engineering Research (IJMER) www.ijmer.com Vol. 2, Issue. 5, Sep.-Oct. 2012 pp-4298-4301 ISSN: 2249-6645

[8] Chetty. Manna Sheela Rani ,” Block based Image fusion technique using Lifting wavelet transform and Neural networks on Medical Images” Research scholar, Dept. of Computer Science & Engg Acharya Nagarjuna University Guntur, Andhra Pradesh, India. International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol. 2, No.5, October 2012

[9] Kanagasabapathy and Vasuki, “Image Fusion Based On Wavelettransform” Department Of Ece, Kumaraguru College of Technology, Coimbatore. International Journal of Biomedical Signal Processing, volume: 2, issue :1, april 2011.

[10] C.M.Sheela Rani, V.VijayaKumar, B.Sujatha “An Efficient Block based Feature level Image fusion technique using Wavelet transform with Neural

network", International Journal of Computer Applications (0975 – 8887), 13-19, Volume 52– No.12, August 2012.

[11] C.M.Sheela Rani, P.S.V.Srinivasa Rao, V.VijayaKumar, "Improved Block based Feature level Image fusion technique using Contourlet transform with Neural network", Signal & Image Processing : An International Journal (SIPIJ), 203-214, Vol.3, No.4, August 2012.