Tracking the Human Motion from Multiple Video Images

Baby Htun, Myint Myint Sein
Research and Development Department, University of Computer Studies, Yangon (UCSY)
babyhtun@gmail.com, myintucsy@gmail.com

Abstract

Object motion and human motion tracking is an important technique used in many systems, especially in the fields of computer vision. An automatic motion tracking and catching system is developed in this research. It can be used to entrap snatchers for shopping mall, office, housing, corridors, borders, secured buildings, at airport and so on. The multiple perspective imagery is applied to track the desired people successfully. It needs to establish the perspective Gaussian Mixture Model and Particle Filter calculation correspondence among the multiple cameras. At once a man pass and is marked or chosen to track any view of camera which is implemented, the system will track the motion of this subject and as well as the other cameras will track him. In this paper, we can track and recognize the subject we desire.

Keywords**:** *Multi-object tracking, tracking-by-detection, image processing, object surveillance, human tracking system, security tracking*

1. Introduction

Human motion analysis is receiving increasing attention from computer vision researchers. This interest is motivated by a wide spectrum of applications, such as athletic performance analysis, surveillance, man-machine interfaces, content-based image storage and retrieval, and video conferencing. And then, it can be used as security checking system.

It can make a precaution by monitoring robbers and strangers from color image sequences obtained from multiple fixed cameras. In particular, we are developing a detection system to track desired subject at sites such as corridors, airports, borders and secured buildings. In our work, the human motion tracking system is developed by security purpose. In human motion and tracking recognition, there may be a gap between perspective two cameras. Tracking work will continue without ambiguous. Multiple motion can tracked all multiple motion objects. But it complex for tracking desired subject among the multiple motion subjects.

Human motion analysis is motivated by applications over a wide spectrum of topics. For example, segmenting the parts of the human body in an image, tracking the movement of joints over an image sequence, and recovering the underlying 3D body structure is particularly useful for analysis of athletic performance as well as medical diagnostics. The capability to automatically monitor human activities using computers in security-sensitive areas such as airports, borders, and building lobbies is of great interest to the police and military. With the development of digital libraries, the ability to automatically interpret video sequences will save tremendous human effort in sorting and retrieving images or video sequences using content-based queries. Other applications include building man-machine user interfaces, video conferencing, etc [1].

In this paper, it gives more favor tracking to cover the security and catching people. This paper focuses mainly on security phase. This contains biometrics (iris, finger print, face recognition) and surveillance-detecting certain suspicious activities or behaviors.

2. Related Works

Q.Cai and J.K Aggarwal Computer and Vision Research Center, Tracking human motion in an indoor environment from sequences of monocular grayscale images obtained from multiple fixed cameras.

Gaussian models are applied to find the most likely matches of human subjects between consecutive frames taken by cameras mounted in various locations.

The system follows the subject moving across the viewing boundary of one camera to another. It is implemented on feature. Weakness is segmentation [1].

Emadeldeen Noureldaim, Mohamed Jedra, Nouredine Zahid, Lab of Conception and Systems, Combine Principle Component Analysis (PCA) and Gaussian Mixture Model (GMM) with Kalman Filter (KF). Attempt to improve tracking of multiple moving objects within the framework of Kalman Filtering [4].

Ben Sigelman is designing to track a human through video sequences with few preconditions. Use a model-based approach based upon Bayesian inference, and approximate the posterior distribution with a particle set. Use raw measurement and dynamic

likelihood mapping. In multiple cameras, tracking in 3D by disambiguating limb locations in the depth dimension for any one specific camera [5].

Sohaib Khan, Omar Javed, Zeeshan Rasheed, Mubarak Shah, are able to discover spatial relationships between the camera fields of view and use this information to correspond between different perspective views of the same person.

They find the limits of field of view (FOV) of a camera as visible in the other cameras. Using this information, when a person is seen in one camera, we are able to predict all the other cameras in which this person will be visible [6].

3. Computer Vision System

Computer vision (image understanding) is a discipline that studies how to reconstruct, interpret and understand a 3D scene from its 2D images in terms of the properties of the structures present in the scene.

The ultimate goal of computer vision is to model and replicate human vision using combines knowledge in computer science, electrical engineering, mathematics, physiology, biology, and cognitive science.

It needs knowledge from all these fields in order to understand and simulate the operation to the human vision system. This paper focuses on computer vision algorithms and their software implementation [2].

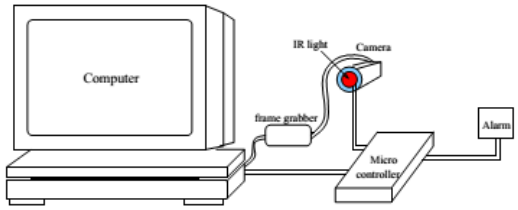

Figure 1. Typical hardware components of a computer vision system

Computer vision system is essential for image processing system. It is a media among human, image and computer system. In computer vision system includes as main types

1. Video input, output and graphics
2. Registration and stereo vision
3. Object detection, motion estimation and tracking
4. Geometric transformations
5. Filters, transforms and enhancements
6. Statistics and morphological operations
7. Code generation and fixed point design
8. Define new system objects. Following is sample code for video file input.

Image processing studies image to image transformation. The input and output of image processing are both images. Typical image processing operations include

- Image compression
- Image restoration
- Image enhancement

4. Tracking Moving Object

When tracking the moving objects, the object may change different poses. So, the system matches the various object locations for example face, back, left side, right side. Objects look different under varying conditions: changes lighting, color, changes in viewing direction and changes in size and shape. In this proposed system, it will use Gradient Cross Correlation Method.

4.1 Perspective Geometric Calculation

It includes computing intrinsic and extrinsic parameters and captures calibration pattern from various views, finally, estimates the moving object position on video sequences [8].

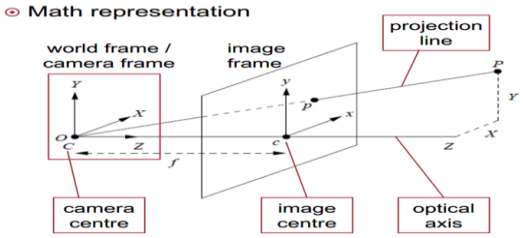

Figure2. Camera Calibration

4.2 Gaussian Mixture Model

In Mixture of Gaussian, each cluster center is augmented by a covariance matrix whose values are re-estimated from the corresponding samples. Instead of using nearest neighbors to associate input samples with cluster centers,

$$P(x|\{\pi_k, \mu_k, \Sigma_k\}) = \sum_k \pi_k \mathcal{N}(x|\mu_k, \Sigma \quad (1)$$

Where are the mixing coefficients, and are the Gaussian means and covariance, and

$$\mathcal{N}(x|\mu_k, \Sigma_k) = \frac{1}{|\Sigma_k|} e^{-d(x,\mu_k;\Sigma_k)} \quad (2)$$

Equation (2) is the normal Gaussian distributions. To iteratively compute maximum likely estimate for the unknown mixture parameters$\{x_k, \mu_k, \Sigma_k\}$, the expectation maximization (EM) algorithm proceeds in two alternating stages:

1. The expectation stage (ES) estimates the responsibilities sample π_k was generated from the k^{th} Gaussian cluster.
2. The maximizing stage (M step) updates the parameter values which are an estimate of the

number of sample points assigned to each cluster.[7][9][10].

4.3 Overview of The Proposed System

In this system, first, the system input is video image files received from three cameras in different positions. And then the system will change these video image files into relevant image frames sequences. The video frame rate is 30 frames per sec. And the image enhancing stage is performed for the segmented images frames. The system let the user can choose or click the desired one subject or more. The calibration parameters of three cameras are calculated. According to the camera calibration parameters, the positions of the selected subject in images frames of three cameras are computed.

Although the selected subject is disappeared in one camera, this subject can be seen in images of others cameras. Finally, the motion of selected subject is tracked through the correlated positions. The system can track if the tracked subject hidden with other things and steadily away from camera views. Figure 3 shows the overview of the proposed system.

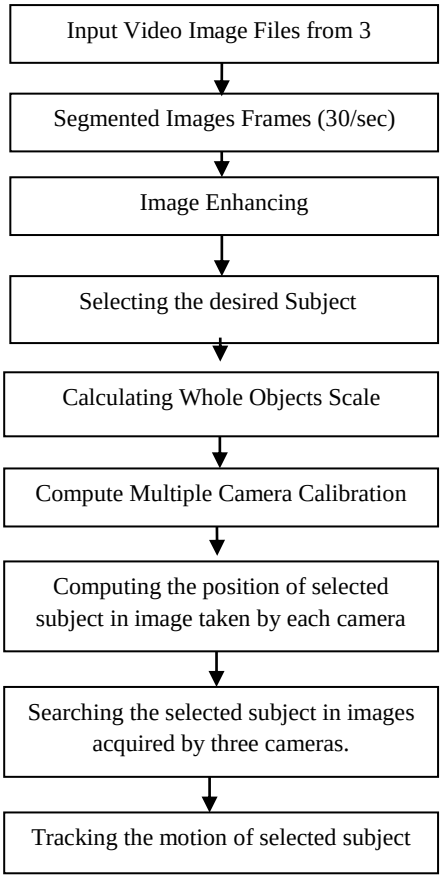

Figure3. System Overview

4.4 Cameras Layout

An automatic motion tracking and recognition system of the desired human subject from multiple camera views is proposed. The proposed system will implement the monitoring system with three cameras.

The desired subjects can be chosen clicking by the system user. If the tracked subject will disappear suddenly from the one camera view, the system can continue to track this hidden subject from another camera views.

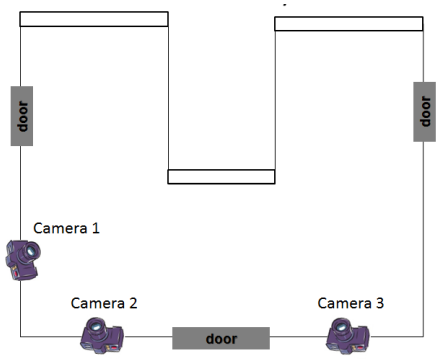

Figure4. Three Cameras Layout

Multiple perspective geometrical calculation, recognition and tracking of human subject are determined. The camera layout is occurred and targeted from inner to outer, especially portico of some big buildings. The layout of the camera system is illustrated in figure (4).

5. Experimental Results

The system can chase the user desired moving subject from multiple moving subjects. So, the camera one is chasing one persons in the figure 5. The camera two will track the moving subject according to camera one.

Figure5. Tracking images (ascending column) from three cameras (ascending row)

However, here, in this figure, the tracked subject is suddenly disappeared from the vision. And then the camera three will track the moving subject tracking by the previous two cameras.

6. Conclusion and Further Works

In remote monitoring system, tracking and recognition is very complex. And it is very important for human environmental security, especially indoor system. Tracking is complex but recognition is more complex because if it is estimated correctly the position and location of the tracking subject in video sequences the system will get success. We can point out moving target subjects in multiple subjects and watch what movement it has such as whether a snatcher will be. The system can give tracking and recognition moving subjects even if the user clicks on or choose it.

As a future work, the system could be extended 3D model instead of 2D model structure. This system can extend to track more darkness images and faded videos frames by using more powerful algorithms. Moreover, it could be tracked the motion of in addition to human beings also animals such as cats, dogs and so on.

References

[1] J.K. Aggarwal, Q. Cai, "Human Motion Analysis: A Review", Department of Electrical and Computer Engineering, The University of Texas at Austin, Proceeding of the 1997 IEEE.

[2] Dictionary of Computer Vision and Image Processing, Robert Fisher, Ken Dawson-Howe, Andrew Fitzgibbon, Craig Robertson, Emanuele Trucco, Wiley, 2005.

[3] Q.Cai and J.K Aggarwal, "Tracking Human Motion Using Multiple Cameras" Computer and Vision Research Center, Detp of Electrical and Computer Engineering, University of Texas at Austin

[4] Emadeldeen Noureldaim, Mohamed Jedra, Nouredine Zahid, "Multiple Tracking of Moving Objects with Kalman Filtering and PCA-GMM Method", Lab of Conception and Systems, Faculty of Sciences, Mohamed V University, Rabat, Morocco 2013

[5] Ben Sigelman, "Video-Based Tracking of 3D Human Motion Using Multiple Cameras", Dep of Computer Science at Brown University, May 2003

[6] Sohaib Khan, Omar Javed, Zeeshan Rasheed, Mubarak Shah, "Human Tracking in Multiple Cameras", Computer Vision Lab, School of Electrical Engineering and Computer Science, University of Central Florida Orlando, FL 32816, 2006.

[7] N.A. Campbell and X. Wu, "Gradient Cross Correlation for Sub-pixel Matching", CSIRO Mathematical and Information Sciences, 65 Brockway Road, Floreat, WA, 6014, Australia, xiaoliand.wu@csiro.au

[8] Ezio Malis, "Camera Geometry", National Institute of Research and Automatic Information Center, Sophia-Antipolis, France, http://www.sop.inria.fr/icare/personnel/mails

[9] Richard Szeliski, "Computer Vision: Algorithms and Applications", Springer, Sep 2010, http://szeliski.org/Book/.

[10] Critian Sminchisescu, Atul Kanaujia, Zhiguo Li, Dimitris Metaxas, "Conditional Models for Contextual Human Motion Recognition", Department of Computer Science, University of Toronto, Canada, USA, International Conference on Computer Vision, ICCV 2005, Beijing, China.