HUMAN MOTION TRACKING AND CATCHING WITH MULTIPLE CAMERAS

[1]BABY HTUN, [2]MYINT MYINT SEIN

[1]University of Computer Studies Yangon (UCSY), Research and Development Department
[2]University of Computer Studies, Yangon (UCSY)
E-mail: 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. The multiple perspective imagery is applied to track the desired people successfully. It needs to establish perspective geometric calculation correspondence among the multiple cameras. The system is implemented by using Gradient Cross Correlation. 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. It can be used to entrap snatchers for shopping mall, office, housing, corridors, borders, secured buildings, at airport and so on. We can study and find about many researches which have various tasks involved in motion analysis of the human body. In this paper, we can track and recognize the subject we desire.

Keywords- Computer Vision System, Cross Correlation, Human Motion Analysis, Monitoring Human Activities.

I. 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 for 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 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. In this paper, it gives more favor tracking to cover the security and catching people.

II. 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.

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.

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. 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.

III. 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.

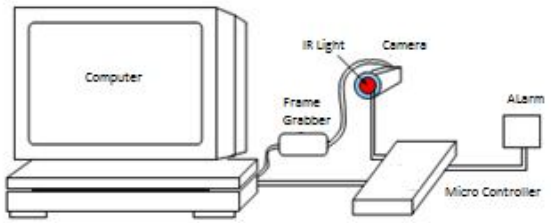


Fig.1 Typical hardware components of a computer vision system

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

Sample code for Video input and output

```
videoFReader=vision. Video File Reader…
('viplanedeparture.avi');
videoPlayer = vision.VideoPlayer;
while ~isDone(videoFReader)
videoFrame = step(videoFReader);
step(videoPlayer, videoFrame);
end
release(videoPlayer);
release(videoFReader);
```

IV. COMPUTER VISION HIERARCHY

- Low-level vision: process image for feature extraction (edge, corner, or optical flow).
- Middle-level vision: object recognition, motion analysis, and 3D reconstruction using features obtained from the low-level vision.
- High-level vision: interpretation of the evolving information provided by the middle level vision as well as directing what middle and low level vision tasks should be performed. Interpretation may include conceptual description of a scene like activity, intention and behavior.
- This paper focuses mainly on middle level and some low level.

V. APPLICATIONS

- Robotics
- Medicine
- Security
- Transportation
- Industrial automation
- Image/video databases
- Human Computer Interface

This paper focuses mainly on security phase. This contains biometrics (iris, finger print, face recognition) and surveillance-detecting certain suspicious activities or behaviors.

VI. MOTION ANALYSIS

There are three types of common motion activities checking system. They are one is motion analysis of the human body structure, two is tracking of human motion using a single or multiple cameras and three is recognizing human activities from image sequences. The relationship among these three areas is widely used as common basic motion techniques. Motion analysis of the human body usually involves the extraction of the low-level feature, such as body part segmentation, joint detection and identification, and the recovery of 3D structure from the 2D projections in an image sequence. Tracking moving individuals using a single or multiple cameras involves applying visual features to detect the presence of humans directly, i.e., without considering the geometric structure of the body parts. Motion information, such as position and velocity incorporated with intensity values, is employed to establish matching between consecutive frames.

After feature correspondence between successive frames is solved, the next step is to understand the behavior of these features throughout the image sequence.

A. 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
- -changes in size and shape

In this proposed system, it will use Gradient Cross Correlation Method.

B. Objectives of This Paper

Followings are objectives of the paper.

- To be able to track someone who want to need tracking.
- To compute and measure the body scale in different views in order to track correctly.
- To compute tracked image positions as geometric feature.
- To achieve multiple tracking if the user requires as possible as.
- To show different colors tracking lines upon different people to see which one is.
- To be able to track sudden appearance of tracked person.
- To know how to track a desired moving subject from multiple moving subjects.

C. Contributions

- The geometrical relations among the multiple cameras calibrations are generated.
- Calculating the object position, skin color and body size (width, length) in order to remember the tracked object.
- To study the role of tracking human motion and recognition analysis.
- Prediction of the human position when he will disappear suddenly from the one camera view, the system can continue to track this defunct subject from another camera views.
- Decide and estimate the tracking line on the subject in view of current camera (c3) whether the object tracked in previous camera (c1) is or not if it was the object in suddenly disappearance in an adjacent camera (c2).

D. 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.

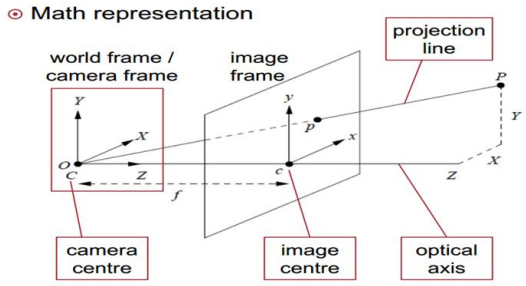


Fig.2. Camera Calibration

E. Gradient Cross Correlation

Image matching should allow for offsets in the target image, and for scaling and rotation. Offsets allow for sub-pixel shifts in the two images, while scaling is necessary when matching images from different sensors or images taken from different distances using the same camera.

Rotation allows for the matching between rectified and un-rectified images or images taken from different viewpoints. The standard approach for area matching for two images to the nearest pixel maximizes the cross-correlation coefficient when the second image is shifted systematically relative to the first over a regular grid.

The formulation of the cross correlation coefficients as:

$$R = \frac{\sum(g_1 - \bar{g}_1)(g_2 - \bar{g}_2)}{\sqrt{\sum(g_1 - \bar{g}_1)^2 \sum(g_2 - \bar{g}_2)^2}} = \frac{C_{12}}{\sqrt{C_{11}C_{22}}}$$

where g_1, g_2 denote the two image intensity values, $\bar{g}_1, \bar{g}_2$ denote the two image average intensity values within the two image patches, C_{11} C_{22} C_{12} denote the two image variances and covariance, respectively.

An affine transformation to calculate the line and pixel in the second image as a function of six parameters can be written as:

$$\begin{cases} x = x_0 + a + Sx.\cos(Rx).y \\ y = y_0 + b - Sy.\sin(Ry).y \end{cases}$$

where x_0, y_0 denote the pixel and line coordinates for the best whole-pixel match on the second image; a, b denote the pixel and line offset or shift; Sx, Sy denote the pixel and line scaling; Rx, Ry denote the pixel and line rotation angles.

VII. PROPOSED SYSTEM

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. Multiple perspective geometrical calculation, recognition and tracking of human subject are determined.

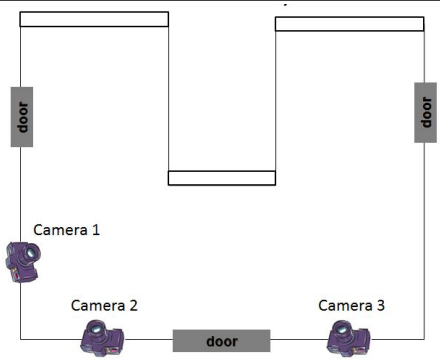


Fig.3. Three Cameras Layout

VIII. SYSTEM OVERVIEW

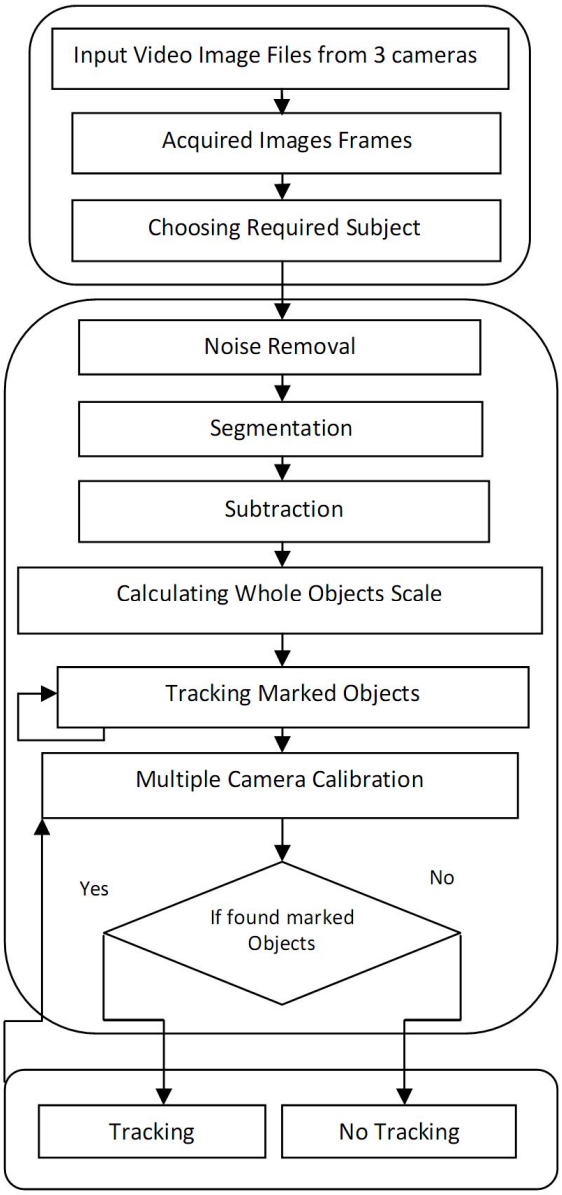


Fig.4 System Overview

In this fig. 4, there are three rounded rectangles. First is system input, second is system processing and three is system output. In this system, first, the system input is video image files received from three cameras. And then the system will change these video image files into relevant image frames sequences. The system let the user can choose or click the desired one subject or more. And they system will track and catch this subject. The system removes the noisy, and segments and extracts foreground from background. The system can track if the tracked subject hidden with other things and steadily away from camera views.

IX. FURTHER WORKS

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.

CONCLUSION

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.

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.

★★★