127 S. K. Kang, J. K. Paik, B. R. Abidi, D. Shelton, M. Mitckes, and M. A. Abidi, "Gate-to-gate automated video tracking and location, " The FAA's 3rd International Aviation Security Technology Symposium, Atlantic City, NJ, November 2001.

Gate-to-gate automated video tracking and location

Sangkyu Kang*, Joonki Paik**, Besma R. Abidi***, David Shelton, Mark Mitckes, and Mongi A. Abidi

IRIS Lab, Department of Electrical & Computer Engineering
University of Tennessee
Knoxville, TN 37996-2100
sangkyu@iristown.engr.utk.edu*, jpaik@utk.edu**, besma@utk.edu***

Abstract

An automated gate-to-gate, airport, video tracking system is proposed and simulated using a virtual three-dimensional (3D) scale model of the airport's main concourse. The proposed research interactively combines two state-of-the-art technologies: (i) video analysis and processing and (ii) 3D visualization and simulation. The video tracking algorithm, which is realized by technology (i), has been developed with emphasis on seamless handover of a moving object using multiple cameras. The simulation environment was created based on Knoxville's McGhee-Tyson airport using Envision software by Deneb.

1. INTRODUCTION

With the number of flights out of almost every airport increasing daily, security has to be of great concern to airport management. The most unpredictable element in any airport terminal is the number of people walking through the halls and passageways. Tracking one person out of a crowd as the person moves from one camera to another is a daunting task for even an experienced operator. The situation becomes even worse if *bolting* occurs when someone, for whatever reason, runs from a security station or security personnel. A *bolter* initiates shut down of the corresponding area, resulting in extensive delays and inconvenience to many travelers. Notable bolting accidents have been reported in [1].

The conventional video surveillance system, Digital Video Recording (DVR) mostly used in airports, focuses on storing image sequences with efficient compression ratios and transferring image sequences via network lines [2,3]. Some of the systems adopt a motion-detection technology to detect a scene change or moving objects, and store only the scene that has motion changes. Unfortunately, this only guarantees a longer recording time, and provides no automatic tracking functions.

More recently, many tracking algorithms have been developed [4,5,6,7]; these algorithms provide good results under certain restrictions, such as images from a stationary camera with a long training time (typically 30 seconds). The most promising result has been provided by Haritaoglu et al. [5]. In this work background images were built using a statistical model for each pixel, which is then used to detect moving pixels. The initial training time, however, is 20-40 seconds, which is too long to update the initial background image when the stationary camera has been moved slightly for whatever reason.

In this paper a new background generation algorithm that is much faster than the Gaussian-based algorithm [5] is proposed, and a simulated Knoxville, McGhee-Tyson airport model with camera handover technology introduced.

2. VIDEO ANALYSIS AND PROCESSING

Generally, an object tracking algorithm is composed of three main functions: background modeling, moving object detection, and object tracking. In this section these functions are described and the object handover technique is also discussed.

2.1 Background Modeling

Moving objects can be identified by the difference between the background and the current images. The simplest means is by using an image without any motion, but this method cannot deal with the problem of changing illumination. Haritaoglu et al. [5] computed the median value over approximately 20-40 seconds, and the standard deviation of the median value is used to generate a background model.

This method provided considerably improved result for extracting moving parts, but the long training time (20-40 seconds) is a disadvantage.

We propose, as an alternative, a new background generation method that requires a shorter training time. Figure 1 depicts this algorithm. In this method, a pixel that does not change for N consecutive frames is used to update the background image.

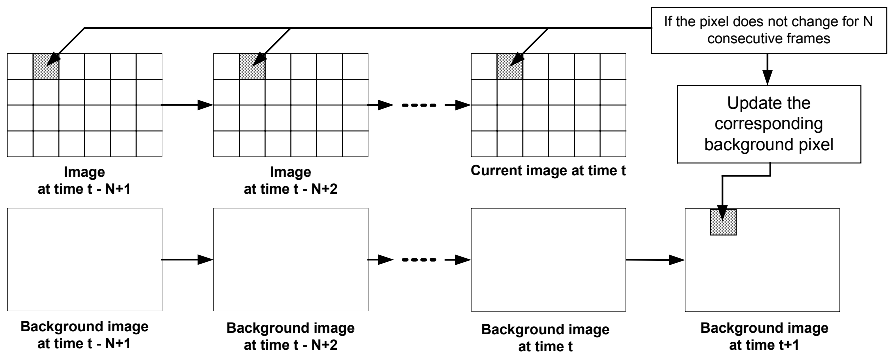


Figure 1: The proposed background generation algorithm.

The procedure implemented for background generation is summarized as follows.

1. If an image is the first image, set the first image as the initial background image.
2. Otherwise repeat the following steps for all pixels.

    2.1 If $|I_k(x,y) - B_k(x,y)| < T_{BG}$: $I_k(x,y)$ is static.

    if $(Flag_{BG}(x,y) > 0)$, then $Flag_{BG}(x,y) = Flag_{BG}(x,y) - 1$

    2.2 else : $I_k(x,y)$ is not static.

    $Flag_{BG}(x,y) = Flag_{BG}(x,y) + 1$

    if ($|I_k(x,y) - I_{k-1}(x,y)| < T_{BG}$ and $|I_k(x,y) - I_{k-2}(x,y)| < T_{BG}$)

    if ($Flag_{BG}(x,y) \geq N$), then $B_k(x,y) = I_k(x,y)$, $Flag_{BG}(x,y) = 0$

    else

    if $(Flag_{BG}(x,y) > 0)$, then $Flag_{BG}(x,y) = Flag_{BG}(x,y) - 1$

$I_k(x,y)$ and $B_k(x,y)$ represent the (x,y)-*th* pixel for the input image and the background image, respectively, at time $t = k$. *N* is the number of consecutive frames as shown in Figure 1, and $Flag_{BG}(x,y)$ is the current number of consecutive frames for (x,y)-*th* pixel. The variable *N* typically depends on a frame rate of input sequences, and we used $N = 5$ for 5 frames/sec video. All operations are performed in the gray-scale pixel intensity domain.

2.2 Moving Object Detection and Tracking

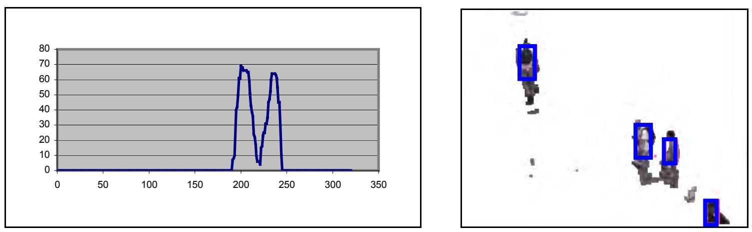
After a background image is generated, moving pixels can be extracted by comparing the background to the current images. $2T_{BG}$, which is determined experimentally, is used as the threshold value to extract moving pixels, and morphological operations, such as opening and closing, are used to remove noise when detecting moving pixels. Labeling follows the morphological operations.

Figure 2(a) shows an input image at the 246th frame and Figure 2(b) the corresponding background image. Thresholded and labeled images are shown in Figures 2(c) and 2(d), respectively.

(a) The 246th frame (b) The background image (c) The thresholded image (d) The labeled image

Figure 2: An example of background generation, thresholding, and labeling processes.

As shown in Figure 2(d), two objects can be considered as one segment due to noise or shadows. By analyzing the histogram of the vertical projections, the segment can be divided. Figure 3(a) shows the vertical projection histogram, using a threshold value of (0.5*maximum), to determine the number of objects. Figure 3(b) shows the resulting segments that have been separated into two objects, each of which is enclosed by a separate rectangle.



(a) (b)

Figure 3: (a) The vertical projection histogram of the largest segment in Figure 2(d), and (b) finally detected object areas enclosed by rectangles.

For tracking purposes, features, such as color distribution, height of objects, and motion information are extracted. Color distribution includes the mean and variance of the U and V components [8] for each segment. Height is an important feature because the height of a human body generally does not change much while the width may vary depending on the position and viewing direction [9]. The final feature, i.e. motion information, includes the direction of change and the acceleration in previous frames for each object area.

In the tracking process, we compute correlation for each feature and determine the best-matched region as the position of the current target object.

2.3 Object Handover Between Cameras

For the seamless tracking by multiple cameras in a spacious, public area, such as an airport concourse, the proposed system should switch the current camera to another that covers a better viewing area. An important criterion in determining a better viewing direction is the distance between the object being tracked and each camera. Generally speaking, a closer camera provides a better image of the object, if it includes the whole object with minimal occlusion.

Two features, the height ratio and motion information, can be used to determine the best viewing camera. For different cameras, the captured images with the same object may have different object sizes because the distances between the cameras and the object are different. The height ratio can be computed by considering actual positions and zooming ratios for each camera. For this research, a simulated environment based on Knoxville's McGhee-Tyson airport was used for determining the height ratio.

In section 4, experimental results using two cameras that capture the images in the same direction and cover the same model are presented.

3. 3D VISUALIZATION AND SIMULATION

An excellent testing field for the tracking algorithm and camera handover strategy would be an airport where multiple cameras could be installed. It is, however, impossible to install such a tracking system for evaluation purposes because every airport has its own security system which might be interrupted by this experiment. An alternative test method is to build a virtual 3D scene model. A 3D scene model of an airport environment is a priceless tool for various testing procedures, especially for object handover among multiple cameras. For example, when a tracked object moves to a boundary of the current view, the current camera should hand-over the tracking task to the nearby camera for seamless tracking. With the 3D scene model, this situation can be performed with different camera parameters, such as zoom ratio, pan/tilt speed, and camera position. Envision software by Deneb was used for modeling Knoxville's McGhee-Tyson airport.

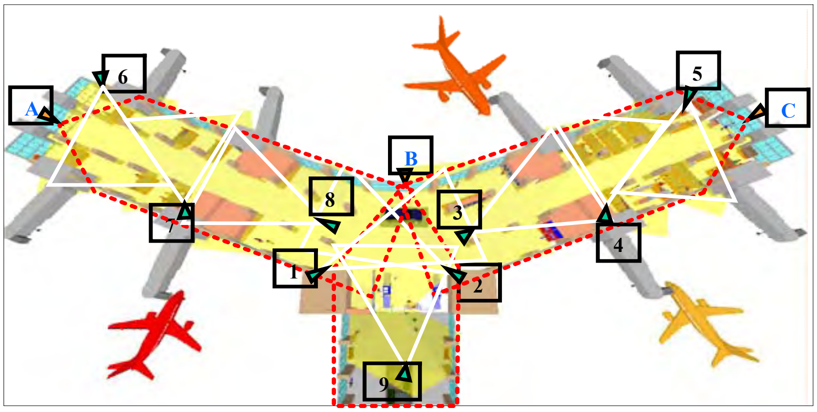


Figure 4: Camera placement scenario with wide-angle cameras (A,B,C) and Pan/Tilt/Zoom cameras (1 to 9). White triangles represent the fields of view of the pan/tilt/zoom cameras in an ordinary situation and the dashed-line pentagons correspond to the areas of coverage of the wide-angle cameras.

Layout of the virtual, 3D airport model is shown in Figure 4. Three wide angle, stationary cameras, which guarantee overall views of the airport, and nine Pan/Tilt/Zoom (PTZ) cameras are assumed to be installed. In ordinary situations, PTZ cameras capture the image of each pre-defined region depicted by the white triangular areas.

5. EXPERIMENTAL RESULTS

Tracking results using two stationary and one PTZ camera are shown in Figure 5. As shown in Figure 5(a), the tracking algorithm was applied to each image from stationary cameras, and the detected object is highlighted with a rectangular box. According to the position of the detected person, the PTZ camera changes parameters (PTZ factors) to get the best view, with the result as shown in the 3rd image. Figures 5(b) and 5(c) show results of the camera hand-over task. When the person disappears in the first viewing area and appears in the second viewing area, the proposed system could successfully hand over the object between cameras.

In 3D scene model shown in Figure 6, the area highlighted by the circle in Figure 6(a) represents the position of an object, and Figures 6(b) and 6(c) show different views covering the person from cameras 2 and 3 as shown in Figure 4. Using this model, the person's height and location and the corresponding PTZ parameters, such as panning angle, tilt angle, and zoom ratio can be verified. Based on this result, we could successfully test the camera handover strategy and perform the virtual reality experiment.

(a) Images at the 18th frame

(b) Images at the 75th frame

(c) Images at the 77th frame

Figure 5: Tracking results using two stationary and one PTZ camera. The first and second columns respectively show images captured by the first and second stationary camera. The third column shows images captured by the PTZ camera.

(a) (b) (c)

Figure 6: 3D scene model of Knoxville's McGee Tyson Airport; (a) top view of the east wing of the concourse, (b) close up view from camera 2, and (c) close up view from camera 3. Camera positions are given in Figure 4.

6. CONCLUSIONS AND FUTURE WORKS

A video tracking algorithm based on updating the background and camera hand-over technologies was proposed. Experimental results show acceptable object handover when a person walks to the next camera's viewing area, and a PTZ camera can continuously follow the person to automatically guarantee the best views. A 3D scene model of Knoxville's McGhee-Tyson airport was also built to minimize the cost and time of implementing the proposed system in an actual airport. Experimentation guaranteed feasibility of the model in testing the camera handover strategy. Future work will focus on tracking tasks at the real McGhee-Tyson airport using multiple cameras.

ACKNOWLEDGEMENTS

This project was supported by The National Safe Skies Alliance Project #R01-1344-49 through Federal Aviation Administration grant #200-G-019.

REFERENCES

[1] Wired News, Flight of Fancy, http://www.wired.com/news/topstories/0,1287,21463.html.
[2] Neurodynamics, http://www.neurodynamics.com.
[3] Nice Security, http://www.nice.com/security/.
[4] U. Neumann, S. You, "Integration of region tracking and optical flow for image motion estimation," Proc. 1998 International Conference on Image Processing , vol.3, pp. 658 –662, 1998.
[5] I. Haritaoglu, D. Harwood, and L. Davis, "W4: Real-time surveillance of people and their activities," IEEE Trans. Pattern Analysis, Machine Intelligence, vol. 22, no. 8, pp. 809-830, August 2000.
[6] L. Lee, R. Romano, and G. Stein, "Monitoring activities from multiple video streams: Establishing a common coordinate frame," IEEE trans. Pattern Analysis, Machine Intelligence, vol. 22, no. 8, pp. 758-767, August 2000.
[7] Q. Cai and J. K. Aggarwal, "Tracking human motion in structured environments using a distributed-camera system," Journal of Pattern Analysis and Machine Intelligence, vol. 2111, pp. 1241-1247, November 1999.
[8] H. Levkowitz, Color Theory and Modeling for Computer Graphics, Visualization and Multimedia Applications, Kluwer Academic Publishers, USA, 1997.
[9] M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis, and Machine Vision, Brooks/Cole Publishing, USA, 1999.