A MULTI-CAMERA SURVEILLANCE SYSTEM THAT ESTIMATES QUALITY-OF-VIEW MEASUREMENT

Changsong Shen, Chris Zhang and Sidney Fels

Department of Electrical and Computer Engineering
The University of British Columbia, Vancouver, BC V6T 1Z4
{csshen, chriszc, ssfels}@ece.ubc.ca

ABSTRACT

In this paper, we propose a multi-camera video surveillance system with automatic camera selection. A new confidence measure, Quality-Of-View (QOV), is defined to automatically evaluate the camera's view performance for each time instant. This measure takes into account view angle and distance from subjects. By comparing each camera's QOVs, the system can select the most appropriate cameras to perform specific tasks. We also present an approach to determine the minimum number of cameras and their layout in a convex polygonal room under specific QOV constraints. Finally, we implement an experimental surveillance system, to confirm the stability of our algorithm and validate the critical underlying concepts of QOV.

Index Terms— QOV, Quality-Of-View, multi-camera, camera selection

1. INTRODUCTION

Multi-camera video surveillance systems have generated growing interest recently, because systems relying on a single video camera tend to restrict both the visual coverage and the total resolution available, which usually imposes undesirable constraints. For example, in [5], the subject is required to be isolated. There have been a significant number of recent advances in detecting and tracking people using multi-camera systems. In [3], multiple synchronized surveillance cameras are employed to locate and track the people in a cluttered scene. In [1], a Bayesian net fuses independent observations from multiple cameras by iteratively resolving independency relationships and confidence levels within the graph, thereby producing the most likely vector of 3-D state estimates given the available data. A problem unique to multi-camera systems is the need to decide which camera or subset of available cameras to use in a given instant of time (since many cameras' views will be superfluous). Various measurements have been introduced to qualify the quality of each camera's viewpoints. In [4], a method is proposed for measuring the ambiguity of 2D measurements provided by each view, and then the ambiguity measurement is used for selecting camera for the most accurate match and tracking. In [2], camera selections are based on the visibility of a part and the observability of its predicted motion from a certain camera. However, in most available papers, the view performance of the cameras is not explicitly considered. In a video surveillance system that employs multiple cameras, one key problem is selecting the most appropriate camera or set of cameras with better view performance to perform a certain task arises. It is desirable that the system can automatically pick the right camera or set of cameras with the best view. This choice would aid the selection of data to be analyzed (and to consequently trigger event alarms); moreover, redundant video data from other cameras can be discarded to save computational resources.

In this paper, we define QOV as a metric to measure the view quality of each camera to a given subject, and we describe how we calculate the Quality-Of-View (QOV) of the cameras to a subject so that our surveillance algorithms can select for the best one. The QOVs are compared to determine which cameras are most informative and therefore selected to perform a specific video surveillance task. In addition, we have developed a method to decide the minimum number of cameras and how to layout these cameras under given QOV constraints. This paper is organized as follows: Section 2 provides a brief overview of system architecture. Section 3 presents the definition of Quality-Of-View, including how to detect the view angle and distance between subjects and cameras. Section 4 explains how to decide the minimum number of cameras and how to layout those camera under given QOV constraints. Section 5 summarizes the papers and discusses the future work.

2. SYSTEM ARCHITECTURE OVERVIEW

The system proposed in this paper is outlined in Figure 1. Initially, we use the method in [6] to calibrate cameras. A set of virtual 3D points is made by waving the laser pointer through the working volume. Its projections are found with sub-pixel precision and verified by a robust RANSAC analysis. After calibration, we record a sequence of background images without a person in the scene. For each pixel in the background, we calculate the mean and variance of pixel intensity, resulting in a Gaussian distribution. To determine whether a pixel is in the foreground or a part of the background, its intensity is fit into the Gaussian model of the corresponding pixels. If image pixels are classified as background pixels, then these pixels are used to update background models. Instead of using RGB color space, we use YUV color space in our system to reduce shadows.

Next, automatic camera selection is processed based on QOV calculations, which includes three primitive sub-tasks (occlusion check, distance detection, and view angle estimation). Finally, tracking is applied on the most informative camera.

In the following sections, we describe the details about each sub-task and how QOV is measured.

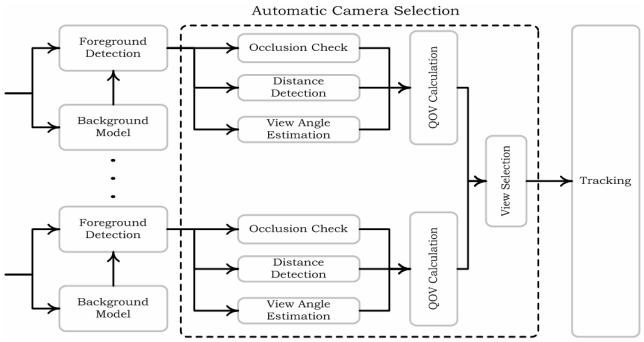


Figure 1 Surveillance System Architecture

3. QUALITY-OF-VIEW MEASUREMENT

3.1. Quality-Of-View (QOV) Definition

In most applications, considering video data from all cameras is unnecessary because video data from some cameras are less informative and sometimes even misleading. Figure 2 shows the same subject's images captured from various view directions. The middle image in the bottom row is captured from the subject's front view, and is more informative compared to other images for many applications, such as face tracking. The corresponding cameras of images in the top row obviously perform poorly in this case. This uninformative video data may trigger wrong alarm, and should not be considered in the system.

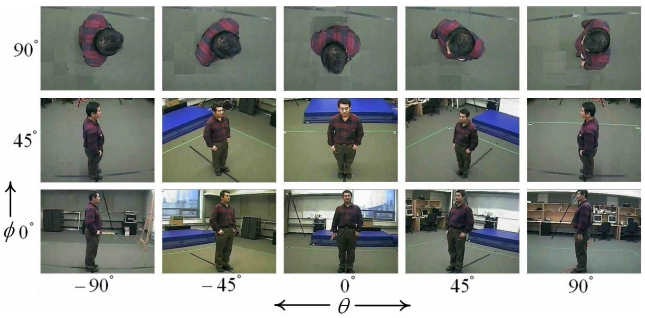


Figure 2 Images Captured from Various View Angles

A confidence measure is therefore needed to address the problem of automatic camera selection. This measure can be used to select the most appropriate set of cameras. It should be an effective indicator of how well the subject is detected. We propose a new measure called Quality-Of-View (QOV).

To arrive at our QOV definition, we need to carefully define two key concepts: view angle and best distance. In our experimental system, human is modeled as an elliptic pillar. We assume the center axis of the pillar is perpendicular to the ground, as shown in Figure 3.

Definition I *Picking a point o and three unit vectors i, j, and k orthogonal to each other defines an orthonormal coordinate frame as (o, i, j, k). We restrict our attention to right-handed coordinate system. The point o is at the COG (Center of Gravity) of the subject's head, i is parallel to subject's body orientation, and k is perpendicular to the ground. Then the View Angle is defined as* (θ, ϕ)*, where θ is the angle between subject's body orientation and the projection of camera optical axis on plane(o, i, j), and* ϕ *is the angle between line passing through the camera center and COG of subject's head and plane (o, i, j), as shown in* Figure 3.

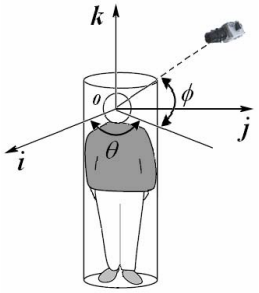


Figure 3 View Angle Definition

Figure 2 shows how (θ, ϕ) affects the image quality. Images, with view angle θ from -90^0 to 90^0, and ϕ from 0^0 to 90^0, are shown.

Definition II *Given a camera and subject, when the height of the subject equals to the height of image, the distance between camera and subject's COG point is defined as Best Distance* L_B.

Definition III *Given a camera* $i \in \{1,...n\}$, *let* $p \in \{0,1\}$ *denote whether or not a subject is out of view or occluded, θ be the orientation angle of the subject body to the camera, and L is the distance between subject and camera, then the Quality-Of-View for that subject is defined as:*

$$QOV = P_i \cap \left(\omega_\theta * \left(1 - \left|\frac{\theta_i}{\pi}\right|\right) + \omega_\phi * \left(1 - \left|\frac{2\phi_i}{\pi}\right|\right) + \omega_l * \left(1 - \frac{L_i}{L_{Bi}}\right)\right) \quad (1)$$

Here, ω is the weight value, which is dependant on specific applications. $\theta_i \in (-\pi, \pi]$, $\phi_i \in [-\frac{\pi}{2}, \frac{\pi}{2}]$

The higher a subject's QOV value for a given camera, the more informative is the corresponding target for that camera, and the more likely that the tracking can be performed yielding accurate results. Thus, for example, tracking can be transferred between cameras based on QOV. For cameras requiring other view angles, the estimation for (θ, ϕ) can be used to define a QOV for better view to select for appropriate cameras.

3.2. Distance Detection and Occlusion Check

The COG and a distance map are obtained by applying distance transforms to the binary result image derived from the background subtraction process. Each value in the distance map corresponds to be the minimum distance to the background. From, these, we calculate the distance between camera and the subject's COG point. Finally we use Best Distance L_B to normalize this distance value. Occlusions are examined by using position detection results.

3.3. View Angle (θ,ϕ) Estimation

First, we explain how to estimate view angle ϕ. During system initialization, we put three bright spots on the floor. Using their locations, we can then determine the ground plane at absolute coordinate. Because ground plane is parallel to the plane *(o, i, j)* in view angle ϕ definition, the angle between the line passing through cameras center and COG of the subject's head and ground plane equals to ϕ. The locations of camera centers in the absolute coordinate system are obtained during camera calibration. Then, from the COG location of the head in the absolute coordinates, the view angle ϕ is estimated, as shown in Figure 4.

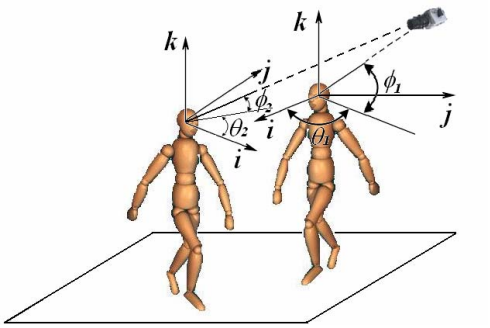


Figure 4 View Angle ϕ Estimation

Next we discuss how to estimate view angle θ. Because the person is modeled as an elliptic pillar in our experimental system, as shown in Figure 5, the person can be describes as follows at *X'Y'* coordinate system () from top view.

$$\frac{x'^2}{a^2}+\frac{y'^2}{b^2}=1 \tag{2}$$

where *a, b* is the semimajor and semiminor axes.

XY coordinate system is actually a rotation of *X'Y'* by view angle θ. Thus, the human model is described by

$$\frac{(x\sin\theta+y\cos\theta)^2}{a^2}+\frac{(x\cos\theta-y\sin\theta)^2}{b^2}=1 \tag{3}$$

Let z denote the distance between subject and camera center, and f denote the distance between camera principal center and image plane, and s denote half the subject projection width. Then we can have

$$s=\frac{f}{2z}\sqrt{a^2\cos^2\theta+b^2\sin^2\theta} \tag{4}$$

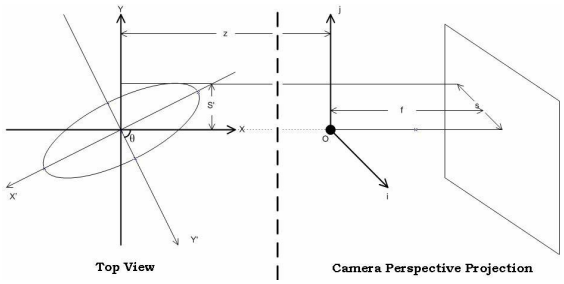


Figure 5 View Angle θ Estimation

Given some sample data set, we can obtain the values of *f*, a and b. Then, from above equation, we can estimate the subject view angle θ.

3.4. Experiment Results

Experiments with real video sequences have been carried out in order to test the performance of the proposed measurement.

Figure 6 Distance Affects View Quality

Figure 6 shows images with various estimated distances between subject and camera from left to right. As distance becomes shorter from left to right, the QOV of the subject increases accordingly.

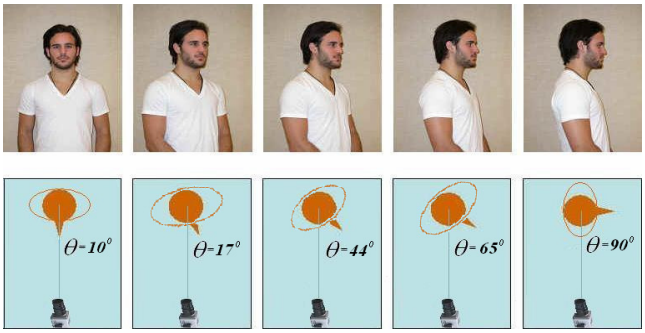


Figure 7 View Angle θ Estimation Results

Figure 8 View Angle ϕ Estimation Results

Figure 7 and 8 shows the estimated view angle (θ,ϕ) value. As θ and ϕ becomes wider from left to right, the QOV of the subject decreases accordingly.

4. CAMERA CONFIGURATION UNDER QOV CONSTRAINTS

4.1. QOV Constraints

Given the measurement of view quality, another question arises: what is the minimum number of cameras required and how should they be configured to ensure at any given instant of time a subject in the scene can be tracked by at least one camera with view angle no worse than some minimum view quality.

We propose the term Coverage Rate (CR) to describe the metric for measuring how well a given camera configuration covers a room. At the moment, we have only defined CR for the 2D case. Suppose camera C_1 covers region R_1 (Figure 9a), R_1 then receives the angular coverage interval 2ψ:

$$A_1=[\bar{d}_1-\psi,\bar{d}_1+\psi] \tag{5}$$

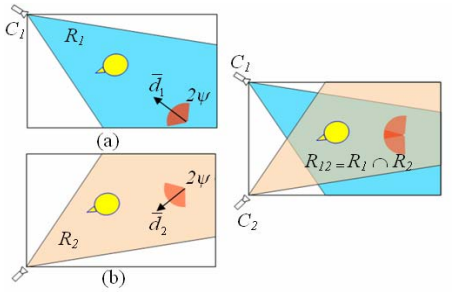


Figure 9 Coverage Rate (CR) Calculation

We can define CR of region R_1 as:

$$Cr_1 = |A_1| / 2\pi = 2\psi / 2\pi \quad (6)$$

For the whole region, we define the CR as:

$$CR = [R_1 Cr_1 + (R - R_1) * 0] / R \quad (7)$$

where R is area of the whole region. Now we add another camera C_2 and suppose the intersection of R_1 and R_2 is R_{12}. Obviously, R_{12} can receive much more angular coverage interval.

$$Cr_{12} = |A_{12}| / 2\pi \quad (8)$$

Formally, CR of the entire polygon is defined as

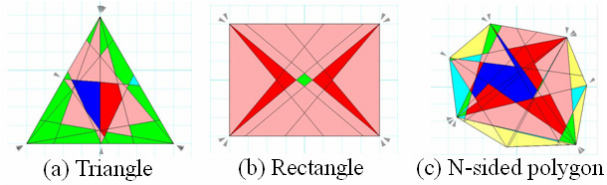
$$CR = \frac{1}{R} \sum_i R_i Cr_i \quad (9)$$

where R_i and R are the areas of the ith sub-polygon and the entire polygon, respectively. The CR's value range is [0, 1].

4.2. Current Camera Layout Approach for 2D

In this paper, our assumption is that cameras can only be placed on the boundary of the room. A greedy algorithm was developed to compute the configuration of cameras to meet the given QOV requirement, while attempting to minimize the number of cameras. Taking a convex polygon as the representative room, our method consists of adding cameras iteratively to the boundary of the polygon so as to marginally increase the largest possible amount of CR per camera. Each time a camera is added to the polygon, it divides the area up into three sets: the sub-polygon to the left of the camera's field of view, the sub-polygon to the right, and set that covered by the camera. These sets consist of the ordered vertices that define the sub-polygon they are meant to represent. In our approach, the first camera partitions the entire environment as above. Each additional camera then acts upon the resulting sub-sets, partitioning each sub-polygon in turn as the first camera had partitioned the entire room. Due to the homogeneity (with respect to coverage) of each set, this is a useful approach, lending well to ease of the computation of the coverage rate. Further, the subsets generated by each partitioning form a minimal cover for the entire room.

As an example, suppose CR goal is 0.8 and all cameras have a fixed FOV of 50˚, Figure 10 shows solutions for a triangle, a rectangle and an N-sided polygon which capture the main complexities of all convex structures. We find the algorithm works well since cameras are scattered uniformly on vertex and edges. For non-convex structures, the solution is to partition them into convex polygons then using above method. However, this will likely result in triangles being formed that require many cameras to get good coverage whereas, if cameras can be placed outside the triangle, a better solution may be possible.



	(a)	(b)	(c)
Camera Number	9	8	12
Color Value of Coverage Rate	0	0.6	0.7	0.8	0.9	0.95	1

Figure 10 Camera Configurations for Various Polygons

5. CONCLUSION AND FUTURE WORK

In this paper, view quality is explicitly considered in a multi-camera surveillance system. Our QOV measure has been defined to facilitate systems select the most appropriate cameras for their applications. The QOV includes view angle estimation and distance detection. In this way, for example, tracking errors due to inappropriate view data can be reduced. Furthermore, we have proposed a solution to determine the minimum number of cameras and their layout in a convex polygon room based on the QOV constraints. Experimental results have demonstrated the effectiveness of estimating view angle and QOV for a single subject.

Future work includes refining view angle estimation to get more accurate results for multiple subjects and extending the camera configuration method from 2D to 3D.

ACKNOWLEDGEMENT

This research is funded by the Natural Sciences and Engineering Research Council (NSERC), Bell University Laboratories and the Institute for Robotics and Intelligent Systems (IRIS).

REFERENCES

[1] S. L. Dockstader and A. Tekalp, "Multiple Camera Tracking of Interacting and Occluded Human Motion," Proceedings of the IEEE, vol. 89, no. 10, pp. 1441-1455, 2001

[2] I. Kakadiaris and D. Metaxas. "Vision-based animation of digital humans." In Computer Animation, pages 144-152. IEEE Computer Society Press, 1998.

[3] A. Mittal and L. S. Davis, "M2tracker: A multi-view approach to segmenting & tracking people in a cluttered scene" International Journal of Computer Vision, 51(3), 2003

[4] E. J. Ong and S. Gong, "Tracking Hybrid 2D-3D Human Models from Multiple Views", in International Workshop on Modeling People at ICCV'99, Corfu, Greece, Sep, 1999

[5] R. Rosales and S. Sclaroff, "Inferring body pose without tracking body parts", in Proceedings of Computer Vision and Pattern Recognition, South Carolina, pp. 721-727, 2000

[6] T. Svoboda, D. Martinec, and T. Pajdla. "A convenient multi-camera self-calibration for virtual environments", Teleoperators and Virtual Environments, pp. 407-422, 2005