AUTOMATIC SCHEDULING OF CCTV CAMERA VIEWS USING A HUMAN-CENTRIC APPROACH

Pradeep K. Atrey, M. Anwar Hossain and Abdulmotaleb El Saddik

Multimedia Communications Research Laboratory
School of Information Technology and Engineering
University of Ottawa, Ottawa, Canada
Email: {patrey, anwar, abed}@mcrlab.uottawa.ca

ABSTRACT

In large scale surveillance systems, a number of CCTV cameras are installed in distributed premises and are connected to a central control station, where human operators observe the different camera views for identifying a probable security breach. In such situations, it is particularly difficult for the operator to pay attention to all camera views. Studies have shown that a human operator can effectively monitor only four camera views at a time. This paper attempts to solve the problem of dynamically selecting and scheduling the four best CCTV views. We adopt a human-centric approach in which the system computes the operator's attention in the CCTV views to automatically determine the importance of events captured by the respective cameras. The experiments show that the proposed method helps a human operator in identifying important events occuring in the environment.

1. INTRODUCTION

In large scale surveillance installations, the CCTV cameras are installed at distributed places in a premise and are connected to a central control station, where a human operator (e.g. a security personnel) remotely monitors the images captured by the cameras.

In the majority of these systems, the number of CCTV cameras are very high (a few hundred), while the human operators who watch the CCTV images are less. A study has shown that the ratio of operators to cameras can be as low as 1:16 [1]. On the other hand, the visual attention of human operator drops below the acceptable level while performing the task of visual monitoring [2]. Besides, it has been found that a human operator can only effectively monitor four CCTV monitors at a time [3]. This practical constraint raises the issue of which four CCTV views, out of several views, should be selected by the system for display at a particular time instant. We assume that the four selected CCTV monitors are displayed adjacent to each other at a higher resolution. As the "importance" of the events captured by the CCTV cameras usually changes over time, the selection of the relevant CCTV views should be performed at a regular interval, which makes the problem of scheduling the CCTV views a challenging one.

In the context of surveillance and monitoring, the importance of an event can be perceived as the degree of deviation of the new observations from the normal happenings in the environment. While, on one hand, the important events can be detected by the traditional computer vision techniques including feature extraction and classification; on the other hand, the importance of an event can also be determined by observing the human operator's attention on the respective CCTV views.

In this paper, we essentially address the problem of scheduling CCTV views for real time monitoring. The core idea of our approach is to adopt a human-centric approach in which the system computes the human operator's attention in the CCTV views to automatically determine the importance of events captured by the respective cameras. Our aim is to reduce the human interaction and to make the monitoring process unobtrusive. To achieve this, the proposed method advocates to use a camera to capture the operator's watching behavior. In particular, the operator's eyes are tracked and his/her attention towards one of the four CCTV views is determined. Based on the attention of the operator, the importance of the CCTV views is computed and then these views are scheduled accordingly for display at a higher resolution.

To determine important views, the existing approaches include change detection in the subsequent video frames assuming that an event would trigger a change. The change-detection is performed by using different techniques such as frame differencing and foreground/background subtraction. Although the change detection method may provide the basis of the initial selection of important CCTV views, it does not always reflect their true importance as this method suffers from several limitations, e.g. poor performance in light changing conditions. Therefore, to overcome these limitations, we propose to follow a *human-centric approach*, which determines the importance of views based on a human operator's monitoring behavior. We envision that the proposed method can assist a human operator and reduce the burden

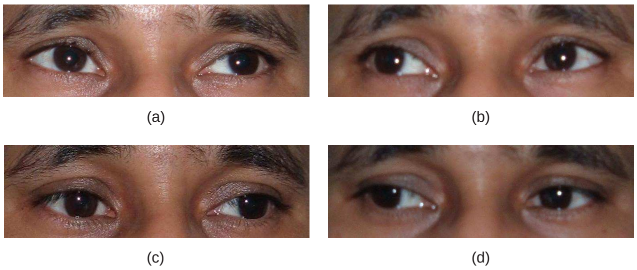

Fig. 1. Four eyes orientations for four CCTV views: (a) top-left (b) top-right (c) bottom-left (d) bottom-right

of monitoring several CCTV views together. Our method has an added advantage of tagging important events stored in a database, which makes the retrieval of these events easier.

2. PROPOSED METHOD

Assuming that, for an operator, the total number of cameras is 16 [1], the proposed method of selecting and scheduling four CCTV views works as follows. Initially, the change detection method is used in each of the CCTV cameras. The four CCTV views with the maximum change are selected for display. If there is no change in any of the views, the four views are selected randomly. Once the four cameras are selected for the display, their importance is computed based on the attention of the human operator. If the operator pays attention to one camera more than the others, it is inferred that the event captured by that camera has a high importance and hence the corresponding CCTV view would have a high importance. The degree of attention is determined based on the duration for which the operator looks at a particular CCTV monitor. The more an operator pays attention to a CCTV view, the higher the importance of that view would be.

In the following, we first describe the preliminary steps of eye tracking and the CCTV view importance computation and then elaborate on the CCTV views scheduling strategy.

2.1. Preliminary Steps

2.1.1. Eye tracking

The objective of tracking the eyes of the human operator is to determine his/her attention in a particular CCTV view. A human operator can have four possible eye orientations when paying attention to the four CCTV views, as shown in figure 1. These are top-left, top-right, bottom-left and bottom-right. Note that the straight eyes orientation is not shown in the figure. In the straight eyes orientation, the operator is assumed to have equal attention to all the four CCTV views; hence this case has been ignored for computing relative attention.

To determine the eyes orientation, a method similar to [4] is adopted. Using this method, the eye's position and its center is detected. Next, we compute the distances d_l, d_r, h_t and h_b (as shown in figure 2) corresponding to the eyes and determine the eyes orientation with the reasoning provided in Table 1.

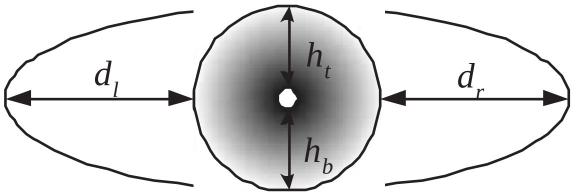

Fig. 2. An eyes sketch with distances d_l, d_r, h_t and h_b.

Table 1. Eyes orientation determination

Reasoning	Eyes orientation
if $d_l/d_r \approx 1$	straight
else if $d_l/d_r < 1$	left
if $h_t \leq h_b$	top-left
else	bottom-left
else if $d_l/d_r > 1$	right
if $h_t \leq h_b$	top-right
else	bottom-right

2.1.2. Computation of CCTV view importance

Once the orientation of the eyes of the human operator is determined, the system observes the duration for which the operator continuously looks at a particular CCTV view. The longer this duration is, the higher the importance the CCTV view would have.

To quantify the amount of importance of a particular CCTV view, a *time-based strategy* has been adopted, which works as follows. Let Δ be the minimum time period for which the four CCTV views remain persistent once displayed. Normally, when the operator looks straight, he/she pays equal attention to all the four views. However, once the operator observes an important event on a particular CCTV view, s/he starts concentrating on it. Let the operator spend $\gamma \leq \Delta$ time in looking straight. The rest of the time $(\Delta - \gamma)$ is used to find the attention of the operator in a particular view. Adopting the strategy that the operator would have attention to a particular CCTV view if s/he concentrates on it for a time period $\delta \geq (\Delta - \gamma)/4$, the importance $I_j \in (0, 1)$ of a particular CCTV view C_j $(1 \leq j \leq 4)$ is computed using the following linear model:

$$I_j = \frac{4 \times \delta_j/(\Delta - \gamma) - 1}{3} \tag{1}$$

In the above model, the bounding values 0 and 1 of I are obtained at $\delta = (\Delta - \gamma)/4$ and $\delta = \Delta - \gamma$, respectively.

2.2. Selection and Scheduling of CCTV Cameras Views

The importance I_j $(1 \leq j \leq 4)$ computed in the previous section is used to schedule the CCTV views. Precisely, the system performs the following steps:

1. For each of the CCTV views, apply a change detection method on the video of p duration. Select four views that have the maximum amount of change. If there is no noticeable change in any of the views, randomly select any four views.

2. Display the images of the selected four CCTV views for Δ duration. During this period, record the operator's watching behavior. For this purpose, the system uses a separate camera. The operator's facial images are processed and the eyes balls are detected. From the position of the eyes ball and its center, the eye's orientation is determined as described in Section 2.1.1. Based on the eye's orientation, the importance I_j $(1 \leq j \leq 4)$ of each of the four CCTV views is computed using equation (1).

3. The CCTV views that have importance $I > 0$ remain persistent. The other CCTV views that have zero importance are replaced after Δ time with the CCTV views (if any) that are adjacent to the important views. In case there is no adjacent view available, the change detection method is used to determine the next CCTV views.

4. Step 3 is continued until any of the CCTV views have importance. If all the views are reset to the zero importance level, Step 1 is followed.

3. RESULTS

3.1. Experimental Setup

To show the utility of our method, we have simulated a CCTV control room environment in which we assumed that 16 CCTV cameras were connected to a central control station and four of these cameras were displayed at a higher resolution. The video frames of 16 distinct places (corridors, laboratory, elevators, stairs and office) in our graduate school were recorded via 16 web cams for six hours. In addition to the normal activities of the people in these places, we requested some of our colleagues to perform special activities such as abandoning a bag in the corridor, doing vandalism near an office door, running in the corridor, etc. The recorded video frames were played on a simulated CCTV panel (an electronic board of $4.5' \times 3'$ size) and a human operator was asked to monitor these images. In our experiment, the value of Δ has been set to 30 seconds.

A separate camera was used to capture the eyes movement of the human operator. The images of this camera were processed for the detection of the eyes ball position and its center. Based on this information, the orientation of the eyes of the operator were determined (using the strategy discussed in Section 2.1.1), and then the proposed steps (described in Section 2.2) were used for selecting and scheduling the four CCTV views.

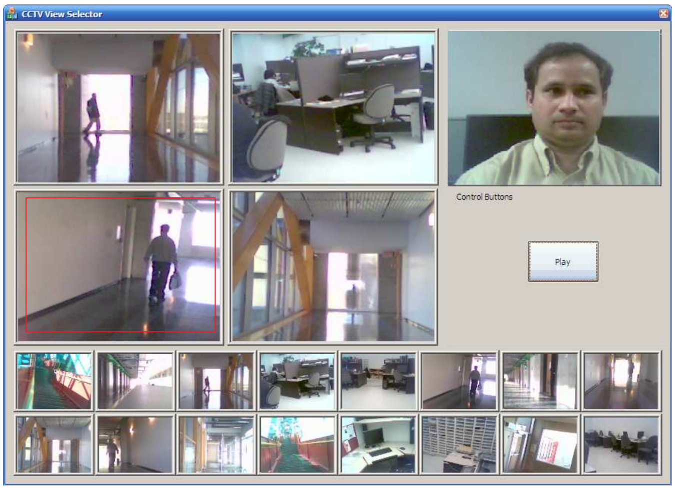

Fig. 3. The GUI of our test bed

The GUI of our test bed is shown in figure 3. It consists of two main parts. The lower part shows 16 small CCTV camera views, whereas the upper part displays - 1) the best four CCTV views selected and scheduled using our proposed method, and 2) the image of the operator showing his eyes movement. Based on the operator's eyes orientation, one of the four CCTV views to which he pays more attention is highlighted with a rectangle on it. The snapshot shown in figure 3 illustrates an instance when the operator is looking at the right-bottom CCTV view.

3.2. Analysis and Discussion

We present the results of the following two test cases. First, we show a specific case when a CCTV view captures the attention of the operator and accordingly the other CCTV views are scheduled over a timeline. Second, we demonstrate that our method of selecting and scheduling the CCTV views helps the human operator to better identify events occurring in the environment.

The first case is illustrated in figure 4. In this figure, the x-axis denotes two specific time instances and the y-axis shows the four best CCTV views selected using our method. For each of the CCTV views, along with the three images, their importance has also been depicted in respective graphs. It can be seen from the figure that, at time instance 1, the operator's attention was moderate in CCTV view 1 and was high in CCTV view 3 as events have been captured in these views. Other CCTV views (2 and 4) did not receive any attention as

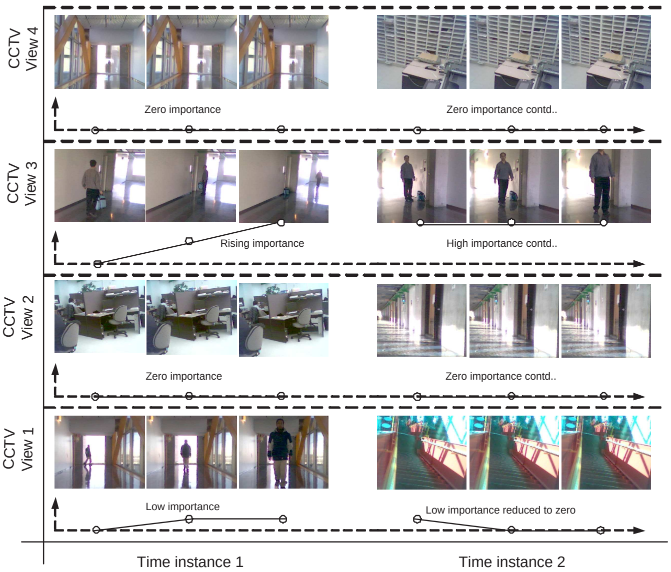

Fig. 4. An illustration of scheduling of the four CCTV views at two consecutive time instances

there were no events. Based on the attention of the operator, the importance of the CCTV views were calculated using equation (1). At time instance 2, the attention of the operator continued in the CCTV view 3 as it showed a suspicious event of a person abandoning a bag in the corridor. Hence, the importance of this CCTV view continued to remain high. On the other hand, the importance of CCTV view 1 is reduced to zero as it showed a normal event of a person walking in the corridor and the operator did not pay much attention to it.

To demonstrate the advantage of our method, we performed the following experiment. An operator, who was knowledgeable of the important events (e.g. a person abandoning a bag in the corridor and/or a person standing near the door of an office for a long time), was asked to monitor the CCTV views in the control room in two different cases. In case 1, all the 16 CCTV views were monitored simultaneously; while in case 2, only the four best CCTV views that were selected and scheduled using our method were monitored. In both the cases, the operator was asked to record important events. While in case 1, the operator recorded the important events by clicking on the CCTV views; in case 2, they were recorded automatically based on the importance value computed using our method. The results of these two cases are shown in Table 2.

We analyzed that, out of 45 important events, 36 events (80%) were identified by the operator when our method was used; while only 18 events (40%) could be identified using the traditional monitoring approach. While using the traditional approach, the operator missed several events as he could not pay attention to all 16 CCTV views together. However, with our method, a set of events were initially displayed on the four CCTV views, which helped the operator in identifying important events. With our method, the operator missed only 20% of events. This is because at a few instances the change-detection method failed to identify the important events as the change in the subsequent frames were found below the threshold. In summary, the results showed that our method effectively helped the operator identify important events.

Table 2. A comparison: Our method vs. traditional approach

Case	Number of important events identified (%)	missed (%)
Our method with the best four CCTV views monitored (View selection through eyes tracking)	80	20
Traditional approach when all 16 CCTV views monitored (View selection by user's clicks)	40	60

4. CONCLUSIONS

This paper has proposed a method based on a human-centric approach, which dynamically selects and schedules the four best CCTV camera views in a surveillance environment. The experiments have shown that the proposed method has helped the operator in identifying the important events. Although the preliminary results are encouraging, it would be interesting to explore how automatic event detection techniques can be integrated to further improve the utility of the proposed approach. For example, automatic detection of the start and the end of an event could be useful in refining the schedule of CCTV views.

5. REFERENCES

[1] H. M. Dee and S. A. Velastin, "How close are we to solving the problem of automated visual surveillance: A review of real-world surveillance, scientific progress and evaluative mechanisms," *Machine Vision and Applications*, 2007, DOI: 10.1007/s00138-007-0077-z.

[2] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, S. Pankanti, A. Senior, C.-F. Shu, and Y. L. Tian, "Smart video surveillance: Exploring the concept of multiscale spatiotemporal tracking," *IEEE Signal Processing Magazine*, pp. 38–51, March 2005.

[3] E. Wallace and C. Diffey, "CCTV control room ergonomics," Tech. Rep., Police Scientific Development Branch, UK Home Office, 1988.

[4] S. Amarnag, R. S. Kumaran, and J. N. Gowdy, "Real time eye tracking for human computer interfaces," in *IEEE International Conference on Multimedia and Expo*, Washington, DC, USA, 2003, pp. 557 – 560.