MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

Eleview: An Active Elevator Monitoring Vision System

Ping Xiao, Maylor K.H. Leung[1] and Kok Cheong Wong
SAS/CE, Nanyang Technological University
Nanyang Ave, S639798, Singapore

Abstract

This paper describes an on-going novel vision application to detect abnormal activities in an elevator. Difference picture technique and human body motion analysis are proposed to accomplish the interpretation tasks. The interpretation makes use of the spatial and temporal features together with the scenario representation. The results obtained are encouraging and further analysis seems fruitful.

1 Introduction

Security issue is an important study to safe guard against intruders, robbers, assailants, terrorists and guerrilla. The elevators which are the target areas of criminal activities are ideal places for automatic security monitoring because an elevator has a restricted scene with limited types of expected activities. Though many elevators have camera installed, the systems are passive, i.e. they only record data and active human interpretations are needed. It would be ideal if the system has built-in intelligence to distinguish normal and abnormal (crime) scenes. If a suspected crime scene is perceived, the system can sound the alarm to draw attention of an operator. Eleview is a system that probes along this direction. With the proposed system, the operator can monitor hundred of elevators simultaneously. Senior citizens, women and children can benefit immensely from the results derived from this research.

The goal of this study is to monitor the scene in an elevator and understand actions that are occurring. Motions inside an elevator can be classified as normal or abnormal according to the analysis of human body motion. Normally the occupants inside an elevator should stay still with little movement until the elevator stops and people get in and out. It is much complicated in the abnormal situations where various movements may occur. The activities can be classified according to different scenario types similar to the works reported in [1, 2]. A scenario is a sequence of expected events. For example, one would expect a murderous scenario to have the scenes of multiple persons entering into the elevator, struggling and finally the phenomena of red blood stain and / or victim falling down. The proposed system would not be fool proof and should not replace a human operator but it can reduce the burden of supervision and false alarming rates.

Section 2 of this paper provides a brief survey on previous works on security study. Section 3 describes the camera setup of Eleview and Section 4 proposes a plan using scenarios to analyze scenes of activities inside elevators. Section 5 concludes this paper.

2 Previous Works

Security issue is an important study. A review of the current state of art using image processing technique can be found in [3]. Majority of the systems were designed to detect intruders and limited body motion analysis was performed. Though many elevators have camera installed, the systems are passive, i.e. they only record data, and human interpretations are needed. A call for a deeper understanding of motion types was advocated by Custance et. al. [4]. The commercial product, Sentinel [5], which used neural network, seemed to be an answer. However, there is little knowledge about how well Sentinel works [3, 5] and there is no reported results to demonstrate its capability in the application of monitoring crime scene in an elevator. A deeper understanding of human body motion [6, 7, 8, 9, 10, 11, 12, 13, 14] (through research) is needed in order to distinguish normal motions from crime actions. This is the focus of Eleview.

3 Camera Setup & Input Data

The data for experimentation was obtained with the help of Singapore Housing and Development Board (SHDB). SHDB has two different setups of camera(s) on the ceiling of an elevator. In one setup, two cameras (C1 & C2) are fixed to shoot down from the left and right (see Fig. 1a). In the other setup, only one camera is used to shoot down from

[1] To whom all correspondences should be addressed (asleung@ntuix.ntu.ac.sg).

the front to the back where a big metal plate (i.e. the mirror) is fixed (see Fig. 1b). The setups are designed to capture the occupants' faces. Another possible setup (in the lab) is to shoot the camera down vertically from the middle of the ceiling. In this study, only the first setup is investigated for the following reasons:

- Most abnormal activities occur only when there are few people (i.e. less than five) inside the elevator. This is an important simplification.
- An ideal setup should capture the faces of the occupants.

To simplify the computation, this study analyzes data from one camera only. The samples from three input sequences of elevator scenes are shown in Fig. 2. Each picture has a resolution of 144 × 180 pixels.

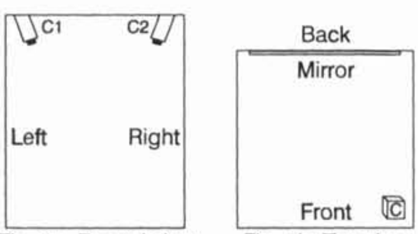

Fig. 1a Frontal view   Fig. 1b Top view

Fig 1. Camera setups inside elevators.

4 Motion Pattern Analysis

The scene inside an elevator can be classified as normal or abnormal. It is expected that little or no motion is detected in the normal cases. On the other hand, abnormal activities can have various movement patterns. The motion types can be classified as:

1. Fighting which creates violent movement.
2. Vandalizing which leaves behind a mess, i.e. overstaying stains.
3. Falling of victims or sick people.
4. Rock&Roll phenomena with activities from shaking bodies to various dancing movements.
5. Reading a book or newspaper.
6. Affectionate activities which include kissing, hugging to a touch on the head.
7. Frisky movements with kids jumping and running around.
8. Still which shows little or no movement.

It is easy for human to distinguish the above abnormal events from normal events but not for the computer vision system. In order to make effective use of the scene information and correctly recognize what is happening, a temporal data structure, Scenario [1, 2], is proposed to represent and recognize scenes in an elevator. A scenario represents an activity as a sequential records of events. Each event can be characterized by the detections of certain features from a scene. Based on the detected features, some scenarios can be triggered to predict future scenes/events to verify the possible scenarios. Because the activities inside an elevator are always restricted and expected, limited number of scenarios can be defined to described all the activities.

4.1 Difference Pictures

Eleview is based on the difference picture method developed by Jain and Nagel [15]. The background picture (f_b), which can be obtained when no one is inside, is subtracted from every frame, $f(t)$, to obtain the masks of occupants, m_o.

$$m_o(x,y,t) = D_o(f(t), f_b)$$

$$= \begin{cases} 0 & if(|f(x,y,t) - f_b(x,y)|) \le T_o \\ 255 & otherwise \end{cases}$$

where D_o is a difference operator and T_o is a threshold which is experimentally set to the value of 12. The mask of the moving regions, m_m, can be obtained by differencing any two consecutive frames as

$$m_m(x,y,t) = D_m(f(t), f(t+1))$$

$$= \begin{cases} 0 & if(|f(x,y,t) - f(x,y,t+1)|) \le T_m \\ 255 & otherwise \end{cases}$$

where D_m is a difference operator and T_m is a threshold which is experimentally set to the values of 12. Additional mask, m_{om}, which records the net translational changes from the previous position, can be obtained by combining the above two as

$$m_{om}(x,y,t) = D_m(D_o(f(t), f_b), D_o(f(t+1), f_b))$$

Conceptually, m_{om} is a subset of m_m which captures the motion changes caused by rotation or translation. These three types of difference pictures capture the motion extents at any moment.

4.2 Proposed Scene Features

Eleview proposes to make use of the following scene features for analysis.

Spatial Features

- *Number of Occupants* : It is assumed that most abnormal activities occur only when there are few people (e.g. less than 5) inside an elevator.
- *Occupant Heights* : This is used to distinguish small playful kids from adults. Small kids tend to create a lot of motion activities.
- *Occupant Bounding Boxes* : The tracking of the number and dimensions of occupant bounding boxes can be used to detect falling and merging of occupants.
- *Occupant Aggressiveness* : Two measures of aggressiveness, i.e. body aggressiveness (AG_b) and fist fight aggressive (AG_f), can be defined as

$$AG_b = \frac{A_i \cap m_{om}}{A_i}$$

$$AG_f = \frac{BC[(A_i - A_{it}) \cap m_{om}]}{A_i}$$

where A_i is the area of the i^{th} occupant, A_{it} is the area of the i^{th} occupant's trunk, and BC is a function to compute the area of the smallest **B**ounding **C**ircle of the swinging arms and kicking legs.

Temporal Features

- *Dwell Time* : This records the time that an occupant stays in an elevator. It is useful to detect occupants that lie down (wounded), or suspicious occupants who refuse to leave the elevator.
- *Duration of Aggressiveness* : This is a good indicator of abnormal activities.

4.3 Scenario Analysis

The scenario is proposed to represent an activity defined by a series of events and the temporal constraints imposed on the events. An event can be considered as a shot of the activity at any instant moment. It is characterized by enabling features which describe the event. Enabling features for an event are required to portray the spatial changes in motion configuration of the event. The temporal conditions may include the occurrence order of events, starting point, duration, tempo and intervals between consecutive events. The occurrence order of events specifies the chronological relationships between events of an activity. Starting point records the start time of occurrence. Duration of an event keeps a record of the minimal and maximal time span. Tempo and intervals then describe the occurring frequency of the events and ranges of time between consecutive events.

Scenario classification is a bi-directional classification process. It includes bottom-up feature collection and top-down verification processes. If the collected evidences match satisfactorily with a scenario, a conclusion is drawn. Generally, the evidences are not convincing and it is necessary to extract more relevant features to guide the analysis. The bottom-up features collection is triggered again to acquire the required features. The bi-directional classification process is effective to decrease search space and save processing time.

5 Conclusion

This paper has described an on-going novel vision application to detect abnormal activities in an elevator. Difference picture technique and human body motion analysis have been proposed to accomplish the interpretation tasks. The spatial and temporal features together with the scenario representation have been employed effectively in analyzing the scenario of an elevator. The results obtained are encouraging and further analysis seems fruitful.

Acknowledgments

We sincerely thank the Singapore Housing & Development Board for its help in obtaining the data for experimentation.

We also thank the Center of Graphics and Imaging Technology, Nanyang Technological University for its support to this project.

References

[1] N.H. Goddard, "Incremental Model-Based Discrimination of Articulated Movement from Motion Features," *IEEE, Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, 1994.

[2] J.K. Tsotsos, J. Mylopoulos, H.D. Covvey, and S.W. Zucker, "A Framework for visual motion understanding, " *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 2, no. 6, pp. 563-573, 1980.

[3] R.M. Rodger, I.J. Grist and G.A.O. Peskett, "Video Motion Detection Systems: A Review for the Nineties," *28th International Carnahan Conference on Security Technology*, Albuquerque, New Mexico, pp. 92-97, 1994.

[4] N.D.E. Custance, K.M. Wickham and K.H. Sage, "Evaluating Scene Monitoring Systems: A Discussion Paper," *International Carnahan Conference on Security Technology: Crime Countermeasures*, pp. 183-189, 1992.

[5] R.F. Mackintosh, "Sentinel: A Technology for Practical Automatic Monitoring of Busy and Complex Scenes," *International Carnahan Conference on Security Technology: Crime Countermeasures*, pp. 190-193, 1992.

[6] M. Herman, "Understanding Body Postures of Human Stick Figures," *PhD Thesis*, Univ. of Maryland, 1979.

[7] J. O'Rourke and N.I. Badler, "Model-based Image Analysis of Human Motion using Constraint Propagation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 2, no. 6, pp. 522-536, 1980.

[8] R.F. Rashid, "Towards a System for the Interpretation of Moving Light Display," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 2, no. 6, pp. 574-581, 1980.

[9] K. Akita, "Image Sequence Analysis of Real World Human Motion," *Pattern Recognition*, Vol. 17, no. 1, pp. 73-83, 1984.

[10] M.K. Leung and Y.H. Yang, "Human Body Motion Segmentation in a Complex Scene," *Pattern Recognition*, Vol. 20, no. 1, pp. 55-64, 1987.

[11] M.K. Leung and Y.H. Yang, "A Region-based Approach for Human Body Motion Analysis," *Pattern Recognition*, Vol. 20, no. 3, pp. 321-339, 1987.

[12] W. Long and Y.H. Yang, "Log-tracker: An attribute-based Approach to Tracking Human Body Motion," *Int'l J. of Pattern Recognition and Artificial Intelligence*, Vol. 5, pp. 439-458, 1991.

[13] K. Rohr, "Towards Model-based Recognition of Human Movements in Image Sequences," *CVGIP: Image Understanding*, Vol. 59, no. 1, pp. 94-115, 1994.

[14] M.K. Leung and Y.H. Yang, "First Sight: A Human Body Outline Labeling System," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 17, no. 4, pp. 359-377, 1995.

[15] R. Jain and H.-H. Nagel, "On the Analysis of Accumulative Difference Pictures from Image Sequences of Real-world Scenes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 1, no. 2, pp. 206-214, 1979.

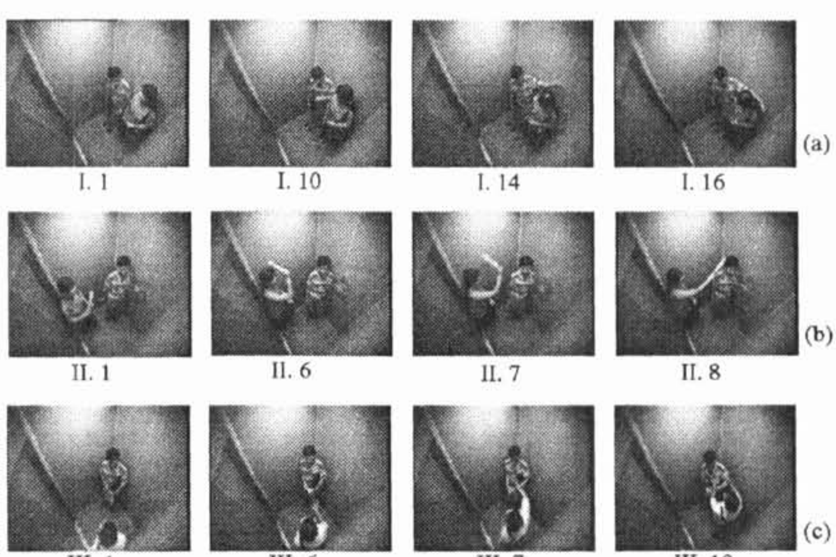

Fig 2. Samples of Input image sequences (Sequences I - III), (a) Sequence I (Normal), (b) Sequence II (Abnormal), (c) Sequence III (Abnormal).