Gaze Focus Detection

Noam Hadas

***Abstract*—This paper was originally submitted to Xinova as a response to a Request for Invention (RFI) on new event monitoring methods. In this paper, a method to utilize the gaze focus of the crowd is proposed.**

In more detail, this describes an idea in which once can locate several video cameras equipped with red flash lights around the perimeter of the event, looking over the crowd. By subtracting subsequent frames taken with the flash firing or not, it is easy to detect the "red eye" reflection, and by knowing which camera took the image know approximately in which direction every individual in the crowd is looking..

I. Abstract

MONITORING a crowd, it may be very useful to get an early indication and warning that anything out of the ordinary is happening inside the crowd mass, and where it is happening, as well as the reaction of the crowd around the source of disturbance.

The paper describes an idea in which once can locate several video cameras equipped with red flash lights around the perimeter of the event, looking over the crowd. By subtracting subsequent frames taken with the flash firing or not, it is easy to detect the "red eye" reflection, and by knowing which camera took the image know approximately in which direction every individual in the crowd is looking. Any disturbance or unusual event in the crowd will therefor present very early as a center point in which direction many people are looking, appearing as the intersection point of many vectors – each starting in an individual, and aimed it his/her gaze direction. The more such vectors intersect at a point, the more sever the warning. The dynamics of such circles forming and evolving around a disturbance may also indicate the crowd reaction – ranging from acceptance to escape efforts.

II. INTRODUCTION

Video surveillance is common practice in managing large scale events involving a large number of participants in a clearly-defined area. Such events may include concert, sport events, political demonstrations and protests and much more. The video images are provided by a large number of video cameras pre-installed at multiple places in and around the event area, and are displayed at the event control and security center. Trained security offices monitor these images trying to identify situations which may develop into a disturbance or anything else that may require action by the event organizers. Unfortunately, detecting suspicious behaviors of an individual or a small group in a large crowd is very difficult, involving just a small fraction of a single image of many. There is there for a need for a software image analysis technology which will be able to identify specific locations in the image which call for enhanced inspection and monitoring.

III. Summary of the invention

Around the perimeter of the event grounds (or around partial areas included in the event) and at higher than human height several video cameras are arranged, all looking towards the crowd. The cameras are equipped with red LED flashes that can emit a very short, high intensity light pulse at a wavelength in which human retina has the highest reflection, i.e. 630mn. Every cycle, two consecutive frames are taken, one with the flash ON, and another with the flash OFF. Subtracting one image from the other results in a dark field, with the eyes of all the people who happen to look in the general direction of the camera appearing as pairs of red dots – any other technology that can detect the eyes of people looking in the direction of the camera, or the direction of the gaze of people in the image will work as well – i.e. IR flashes to facial recognition.

Since multiple cameras are used, each one collects data about people looking in the general direction of that specific camera – or more or less in that direction.

The software generates a vector for each pair of eyes in the image (or face recognized in the image), where the vector starts at the location of the person, the vector is directed towards that camera, and the vector length is proportional to the crowd density at that location since denser crowds limit the distance visible to each individual observer.

All vectors, generated from the data from all cameras, are superimposed on a single, top-view map of the event area.

IV. Detailed Explanation

Normally, most of the people will be looking in one preferred direction – towards the stage, screen, podium or whatever they came to see, at the Front. These vectors will all be parallel and will not intersect at all. Some of the people will look in other directions – at friends, at the area in general, at some interesting decorations and so on, but overall these non-frontal gazes will be few, scattered in the crowd, and transient. Tracing the vectors generated by these gazes will not produce a single, lasting focal point.

On the other hand, if anything out of the ordinary happens inside the crowd, i.e. somebody strips naked, than all the people within a looking distance (in the local crowd density) around him/her will turn their attention to that person by rotating their heads to look. As soon as this happens, all the vectors generated by their gazes form a circle around the disturbance, with the many gaze vectors intersecting at the location of the disturbance. Of course, many such circles may form at the same time at different places due to transient disturbances (i.e. somebody shouting). Still, the very appearance of such focal points that last more than a few seconds is a clear and very early indication of a disturbance, and action can be taken such as pointing a telescopic camera to that location to get a clearer and more detailed image, send a drone or take any other action.

The dynamics of such focal points appearance and evolution may provide a lot of information about their nature and significance. Some possible variations are:

1. Around each disturbance a space where no eyes are detected, growing with time. This means that people close to the disturbance are moving away from it, and this indicates that there may be danger involved – i.e. somebody is waving a knife. Note that the cause of the disturbance may not be visible in the image, but the formation of a vector intersection center and then the clearing of the area around the center from eyes are sufficient indications.

2. The center point forms, than moves leaving a trail of gazes that is growing in length and width with time. This means that the source of the disturbance is moving, with the people who looked at it still tracking its progress. One can expect the trail will grow in width forming a “comet tail” disturbance as more and more people see their neighbors looking in a specific direction and start looking in the same direction.

3. A circle forming and then dissolving in a short time. This means something happened that either stopped, or people lost interest – i.e. somebody brought up a new sign.

4. An small area of “confused” gaze directions, usually growing in size, standing out against a background of more or less unidirectional gaze direction (towards the front). This probably indicates a group fight or unrest which may spread. Usually such an area will start as a small circle around a disturbance, and then the area “absorbs” more and more individuals who turn their gaze from the front to random directions as they engage in the fight.

5. Overall, any small section of the event where the statistical distribution of gaze vectors is markedly different from the surrounding areas is an indication that something is different there, and should be investigated. Following this, it may even be possible to detect a group of individuals with special interests or intentions by the fact that they look more to their leader or engage in some interactions which lead them to look in different directions than the surrounding people at the event.

The process may be better understood by looking at the drawings below.

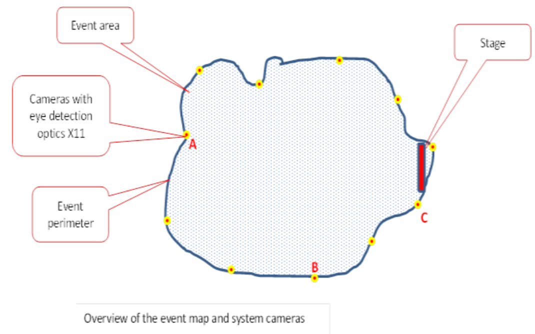

Fig. 1. Overview of the event map and system cameras

Fig. 2. Normal image captured by camera A, note no eyes are visible since everybody is looking to the stage

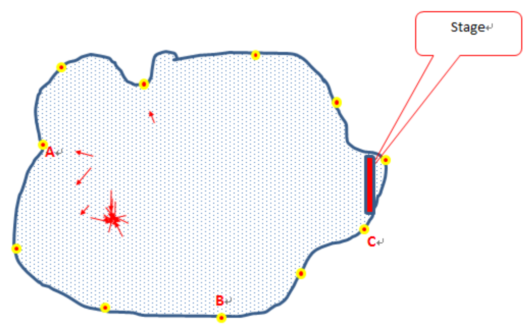

Fig. 3. Event map with disturbance -non-intersecting vectors are not presented for clarity except a few left for demonstration purposes.

In the second drawing the disturbance is easily identified. Some additional non-frontal vectors that are detected by various cameras were added as reference for non-significant vectors. Normally there will be many more such vectors as people do look around occasionally, but they will be dispersed more or less equally over the entire area of the event, they will not be oriented in any specific direction, and they will only deviate from the front direction for short periods. The vectors around a disturbance, on the other hand, will (a) be concentrated around a common center, and (b) be present for much longer durations. Statistic filtering can easily remove most of the non-significant vectors and leave for further analysis only vector sets that may point to a disturbance.

Each vector starts at the location of the pair of eyes that were detected by the camera, and its length may be proportional to the density of people in that area – where the denser the people the shorter the visible range of each person and hence the shorter the vector.

V. Conclusion

Overall, the technology shows a route for providing early detection of any interference in a crowd which draws the attention of the surrounding participants, be it positive or negative, and as such it may be of value to the authorities monitoring the event. Of course, everything described here regarding eye detection will hold just the same for face detection, but this will require much higher resolution cameras, and much more computational power.

The proposed technology is applicable to any public event security management. It has the ability to automatically provide very early warning of risk situations, including their location, severity and development.

Acknowledgment

This research was originally submitted to Xinova, LLC by the author in response to a Request for Invention. It is among several submissions that Xinova has chosen to make available to the wider community. The author wishes to thank Xinova, LLC for their funding support of this research.

More information about Xinova, LLC is available at www.xinova.com.