DETECTING EVENTS IN VIDEO SURVEILLANCE SYSTEM

S.Sivasubramaniam [1] , Dr.M.Ramanan [2]

[1] M.tech

[1] Department of Information Technology ,Coimbatore Institute of Engineering and Technology ,Coimbatore

[2] Department of Physical Science and IT ,Tamilnadu Agricultural University ,Coimbatore,India.

Abstract : Multimedia systems utilize multiple media streams, each of which has different confidence levels in accomplishing various detection tasks. For example, in a multimedia surveillance system, one would usually have higher confidence in an audio stream compared to a video stream for detecting human shouting events. The pre-computation of these confidence levels is cumbersome especially when new media streams are dynamically added to the system. This paper proposes a novel method, which dynamically computes the confidence levels of new streams based on the past history of their agreement/disagreement with the already trusted streams. To demonstrate the utility of the proposed method, we provide the experimental results for detecting events in a multimedia surveillance scenario.

Keywords: Event,Confidence,Surveillance

I. Introduction

MULTIMEDIA systems utilize multiple types of media such as video, audio, text and even RFID for accomplishing various detection tasks. Examples of the detection tasks include:

• determining the occurrence of an event in a multimedia surveillance environment or in a sports video;

• finding the boundaries of a shot in a video;

• detecting a face in an image or a video frame;

• ascertaining the presence of a specific word (e.g., attack) in a recorded audio.

Different types of media possess different capabilities of accomplishing various detection tasks under different contexts. Therefore, we usually have different confidence levels in the evidence obtained based on different media streams for accomplishing various detection tasks.

For example:

• In multimedia surveillance systems, for locating an object, we may have higher confidence in the evidence obtained based on an RFID stream compared to the evidence obtained based on an audio stream. Also, in a dark environment, the evidence about a human's running event, based on an audio stream, could have a higher confidence level than the evidence obtained based on a video stream.

• For the face-detection task, the different cues such as skin color, motion, and sound may have different confidence levels.

We have studied the past works related to the use of confidence in sensors/media streams

2 EXISTING SYSTEM

In the past, confidence has been widely used in the context of data management in sensor networks . Few works have been found in context of multimedia applications. We have studied the past works related to the use of confidence in sensors/media streams from the following four aspects—first, how the confidence level of a stream is determined; second, how the confidence information has been used in the integration of streams; third, how the fusion of confidence values is performed to determine the overall confidence in a group of streams and the fourth, whether or not the confidence has been computed dynamically. The past works have been summarized. While the works determine the confidence in a stream based on how it has helped in obtaining the accurate decisions in the past, the works have used the sensors' current observations in determining their confidence level. Atrey *et al* used a Bayesian formulation, which adopts a logarithmic opinion pool (LOGP) as a consensus rule for performing confidence based assimilation of media streams, but they have considered the static confidence (i.e., pre-computed confidence). The proposed work is different from in that the focus of has been on providing an assimilation framework for event detection by using the pre-computed confidence levels of streams. The main objective of the proposed work is to provide a method for dynamic computation of the confidence levels of streams so that the pre-computation of confidence levels can be avoided. De Silva *et al.* used experimental evidence to provide different weights to audio and video modalities for recognizing human facial expressions. The authors used a linear weighted sum strategy to integrate the different modalities. However, their approach is based on assigning the static weights to audio/video modalities, and they have not provided any method to compute the overall confidence in a group of modalities. In the context of news video analysis,

3 PROPOSED SYSTEM

The system proposes a novel method for dynamically computing the confidence levels of streams in a multimedia system. The proposed method is based on the following idea. Let us follow a trusted news bulletin e.g., the London Times. We also start by following an arbitrary news bulletin and compare the news content provided on both. Over a period of time, our confidence in the arbitrary bulletin will also grow if the news content of both the bulletins have been found similar, and vice versa. Our method computes the confidence levels in newly deployed streams based on the knowledge of the existing

"Trusted" stream(s) and the "Agreement Coefficient" among the newly deployed streams and the existing trusted streams. A stream can be called "trusted" if its confidence level is greater than a threshold. The agreement coefficient between any two streams refers to the degree of how concurring or contradictory the evidences are that the system obtains based on them. The proposed method is generic and can be applied to various detection tasks in multimedia systems where multiple streams (or modalities) of varying capabilities are used. We show its utility for the task of event detection in multimedia surveillance systems

The main objective of the proposed work is to provide a method for dynamic computation of the confidence levels of streams so that the pre-computation of confidence levels can be avoided. De Silva et al. used experimental evidence to provide different weights to audio and video modalities for recognizing human facial expressions. the proposed method is different from all the works described above in that our method computes the confidence level of streams based on their agreement/disagreement with the trusted streams. In fact, we utilize the current observation as well as the past history of agreements/disagreements among streams

4 RELATED WORKS

4.1 VIDEO SOURCE SELECTION

Using samples including digital cameras and scanners, the images are stored and transferred from devices to machine. In the second part of the system considers image sequence identification and Device detection methods that computers could use to achieve better results and process more complicated input.

A video electronic device control signal source selection switch for allowing a user to select the source of signals for controlling the operation of an video/audio electronic device comprising a switch having three orientations for enabling a normal, local, and remote mode of operation.

An input terminal adapted to receive both internal panel control and remote control signals.

And an output terminal for transmitting selected remote and panel control signals, with the normal mode allowing the transmission of both selected remote and panel control signals, whereby permitting the user to control the operation of the electronic device using both the remote control and the panel controls, the local mode allowing transmission of only selected panel control signals, whereby permitting the user to control the operation of the electronic device with only the panel controls, and a transmission mechanism for delivering internal remote and panel control signals from the electronic device to the switch and supplying selected remote and panel control signals from the switch back to the video/audio electronic device for controlling its operation.

4.2 CONNECT TO CAPTURING DEVICE

This article will focus on how to capture video from an analog video source to a Windows XP computer using a

ISSN:-2581-6934 |Available at : www.jidps.com | pp :(21-25)

external Video Capture device. I will show you how, using a standard VCR as the source as the capture device and Pinnacle Studio Plus 9 as the capture software. This how-to would work with any other combination of capture hardware using a USB 2.0 cable, capture software or analog source (such as 8mm, Hi8 or a VHS-C camcorder).

- The user must set up your video capture hardware by plugging in the USB 2.0 cable to the device and connecting it to the port on your PC. Power on the capture device by plugging it into an electrical outlet.
- After turning on your PC. The capture device should be recognized by the PC.
- Connect the source by plugging in the source device's video and audio out cables into the video and audio inputs on the capture device. For a VHS VCR, I am connecting the RCA video (yellow cable) output and RCA audio (white and red cables) outputs to the RCA inputs on the DVD Capture device.
- Start your video capture software. Double click the icon on your desktop or go to Start > Programs > Pinnacle Studio Plus 9 (or the name of the program you're using) to run the software.
- You need to configure the capture software to tell it what format to encode the video to. If you plan on recording to CD, you would pick MPEG-1, for DVD pick MPEG-2. Click the Settings button and then click the Capture Format tab. Change the preset to MPEG and quality setting to high (for DVD).
- To capture your video, click the start capture button and a dialog box pops up for a file name. Enter a file name and click the Start Capture button.
- Once your video is captured to your hard drive it can then be imported into a video editing software application for editing or recorded to CD or DVD using CD/DVD Recording software and a CD/DVD writer.
- The write code to access the driver of the camera by code. Then the properties OS the camera are identified if the driver is installed.

4.3 SURVEILLANCE SYSTEM SETUP

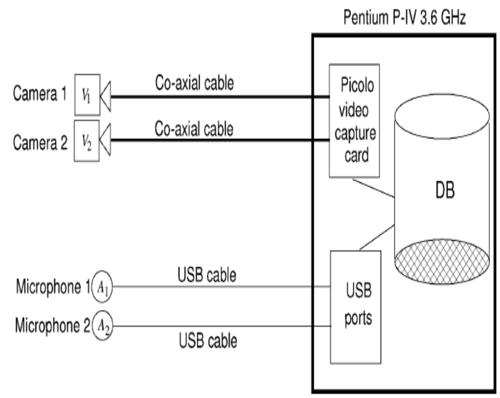

Fig. 2. Surveillance system setup.

The System employed four sensors (1 Canon VC-C50i video cameras and two USB audio microphones) in the corridor of our school building with the objective of surveilling various Usual and unusual activities such as human running, walking, standing, talking, shouting and door knocking. The surveillance environment layout. The two video cameras, (denoted by and) are used to record the video from the two opposite ends of the corridor, and the two audio sensors (denoted by and) are employed to capture the ambient sound. The fields of view of both the cameras. The microphones are unidirectional, and are pointed towards the respective doors of the corridor.

4.4 RECEIVE VIDEO FRAMES

Any camera or video resolution should be always in "landscape mode", usually mirroring the aspect ratios / resolutions of PC screens. to capture frames in exactly 240 x 320 increase the video resolution to 640 x 480, and paint a white 240 x 320 rectangle over the video feed so that you could see frames in the rectangular area. Capture the video at 640 x 480, and the system crops the image to the 240 x 320 rectangle. By adjusting the cam resolution at 320 x 240 the video frames can be extracted, then physically by tilt the webcam by 90 degrees, the system can receive photo at 320 x 240, then rotate the grabbed frame back by 90 degrees. The problem would be that the video feed on the screen would be on its side. But the result is a true 240 x 320 image. A more advanced approach is along the same lines, physically by tilting the webcam 90 degrees..... And write a Framework to pass-through codec

ISSN:-2581-6934 |Available at : www.jidps.com | pp :(21-25)

which turns the webcam video stream from 320 x 240 to 240 x 320 by rotating each frame stored in to the temporary runtime memory

Data Set

The data set for our experiments consist of the recorded video images (of resolution 768 576 pixels) from two video cameras and the recorded audio (of 44.1 kHz) from two USB microphones for more than twelve hours in the corridor of our school building. A total of 92 events occurred over a period of 1375 s during the twelve hours of recording. the details of various events and their time durations.

It is important to note that the events can occur at different granularity in time, location, number of objects, and their activities; and can be categorized as a compound or atomic event .An event is called atomic when an object having one or more attributes is involved in exactly one activity, whereas, a compound event is the composition of two or more different atomic events

Event Detection

1) Based on the Video Stream:

The system detects events of humans standing/walking/running by processing the video frames. The video processing involves background modeling and blob detection. An adaptive Gaussian method is used to model the background. For blob detection, the system first segments the foreground from the background using simple “matching” on the three RGB color channels, and then uses the morphological operations (erode and dilation) to obtain connected components (i.e., blobs). The matching is defined as a pixel value being within 2.5 standard deviations of the distribution. We assume that the blob of an area greater than a threshold corresponds to a human. Once the bounding rectangle for each blob is computed, the middle point of the bottom edge of the bounding rectangle is mapped to the actual ground location using the calibration information of the video cameras.

2) Based on the Audio Stream:

The events such as footsteps, talking, shouting and door knocking are detected using the audio streams. The recorded audio (of 44.1-kHz frequency) is divided into “audio frames” of 50 ms each. The frame size is chosen by experimentally observing that 50 ms is the minimum period during which an event, such as a footstep, can be represented.

CONCLUSION

A method for the confidence evolution in multimedia systems is presented in this paper. The proposed method dynamically computes the confidence level of a media stream based on the past history of whether it helped in obtaining evidence, which is agreeing or contradictory to the already trusted streams in the system. The experimental results provided for event detection in a multimedia surveillance system have shown that the overall accuracy of event detection with the confidence level of streams computed using the two methods (i.e., pre-computed and our method) are quite comparable. Future work would be to explore how the proposed confidence evolution method can be used in other multimodal applications such as biometric authentication etc.

REFERENCES

[1] A. Tavakoli, J. Zhang, and S. H. Son,(Jun.2005) “Group-based event detection in undersea sensor networks,” in *Proc. 2nd Int. Workshop on Networked Sensing Systems*, San Diego, CA.

[2] C. O. Conaire, N. O. Connor, E. Cooke, and A.Smeaton,(Feb.2006)“Detection thresholding using mutualinformation,” in *International Conference on Computer Vision Theory and Applications*, Setubal, Portugal.

[3] D.Yu and D. Frincke,(Mar.2005) “Alert confidence fusion in intrusion detection systems with extended Dampster-Shafer theory,” in *Proc. 43rd ACM Southeast Conf.*, Kennesaw, GA.

[4.] J.Frolik, M. Abdelrahman, and P. Kandasamy,(Dec,2001) “A confidence-based approach to the self-validation fusion and reconstruction of quasi-redundant sensor data,” IEEE Trans. Instrum. Meas., vol. 50, no. 6, pp.1761–1769.

[5]. L. C. De Silva, I. T. Miyasato, and R. Nakatsu,(Sep,1997) “Facial emotion recognition using multi-modal information,” in *Proc. Int. Conf. Information, Communications and Signal Processing*, Singapore,pp. 397–401.

ISSN:-2581-6934 |Available at : www.jidps.com | pp :(21-25)

[6] M. S. Kankanhalli, J. Wang, and R. Jain,(Oct 2006) “Experiential sampling in multimedia systems,” *IEEE Trans. Multimedia*, vol. 8, no. 5, pp.937–946, Oct.

ISSN:-2581-6934 |Available at : www.jidps.com | pp :(21-25)