LINEAR ARRAY OF PHOTODIODES TO TRACK A HUMAN SPEAKER FOR VIDEO RECORDING

D. DeTone, H. Neal, R. Lougheed

UNIVERSITY OF MICHIGAN

ABSTRACT

Communication and collaboration using stored digital media has garnered more interest by many areas of business, government and education in recent years. This is due primarily to improvement in the quality of cameras and speed of computers. An advantage of digital media is that it can serve as an effective alternative when physical interaction is not possible. Video recordings that allow for viewers to discern a presenter's facial features, lips and hand motions have been shown to be more effective than videos that do not. To attain this, one must maintain a video capture where the speaker occupies a significant portion of the captured pixels. However, camera operators are costly, and often do an imperfect job of tracking presenters in unrehearsed situations. This creates motivation for a robust, automated system that directs a video camera to follow a presenter as he or she walks anywhere in the front of a lecture hall or large conference room. Such a system is presented.

The system consists of a commercial, off-the-shelf pan/tilt/zoom (PTZ) color video camera, a necklace of infrared LEDs and a linear photodiode array detector. Electronic output from the photodiode array is processed to generate the location of the LED necklace, which is worn by a human speaker. The computer controls the video camera movements to record video of the speaker. The speaker's vertical position and depth are assumed to remain relatively constant – the video camera is sent only panning (horizontal) movement commands. The LED necklace is flashed at 70Hz at 50% duty cycle to provide noise-filtering capability. The benefit to using a photodiode array versus a standard video camera is its higher frame rate (4kHz vs. 60Hz). The higher frame rate allows for the filtering of noise infrared such as sunlight and indoor lighting – a capability absent from other tracking technologies. The system has been tested in a large lecture hall, and is shown to be effective.

MOTIVATION

The ATLAS Experiment, one of the largest experiments based at CERN, is being built by a collaboration of around 3,000 physicists who are based at institutions spread across the globe. It is one of the largest collaborative efforts ever attempted in the physical sciences [1].

To lessen collaborative difficulty caused by physical separation, busy schedules, and overlapping events, a prototype automated video recording system for meetings, presentations, and lectures was created and is presented here. A system such as this has applications not only for ATLAS scientists but also in industry, governments, and academic institutions.

DESIGN CONSIDERATIONS

Our video recording system is based on infrared (IR) tracking; the person presenting must wear a special necklace with IR LEDs. Apart from being fast and accurate, the robotic tracking and recording system was designed to have the following unique features:

Body Framing – When recording lectures and other presentations, capture of video is essential for viewers to see facial expressions, the speaker's lips, and their hand-motions and demonstrations that are presented. However, camera operators are costly, and often do an imperfect job of tracking presenters in unrehearsed situations. Our system was developed with sufficient accuracy and speed to continuously track the presenter's movements to keep his/her body framed in the video capture.

Fig 1a – An image of a speaker recording system without body framing.

vs.

Fig 1b – An image of a speaker recording system with body framing.

Robust (Noise Filtering) – To be employed on a large scale, the speaker tracking system must work in many environments. Since the system relies on IR to find the target, the system must be able to filter out noise IR sources such as incandescent lights, infrared communication ports found on laptops and assisted listening systems, stairway lights, and sunlight. To accomplish this, the IR LED necklace is flashed at 70Hz.

Fig 2a - Sunlight reflecting off of moving leaves.

Fig 2b - Stairway lighting.

SYSTEM OVERVIEW

Infrared LED Necklace Segment:
The speaker to be tracked wears a special IR LED necklace, as pictured in Fig 3a and Fig 3b. To simplify the fabrication for the proof of concept, eight vertically aligned LEDs were used for testing the linear array sensor instead of a more conventional circular arrangement which would allow for full rotation of the speaker.

Fig 3a – A speaker wearing the prototype necklace.

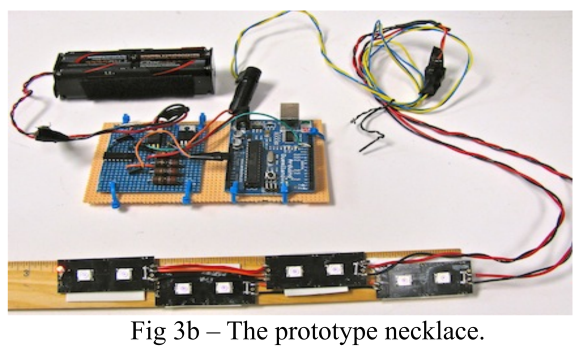

Fig 3b – The prototype necklace.

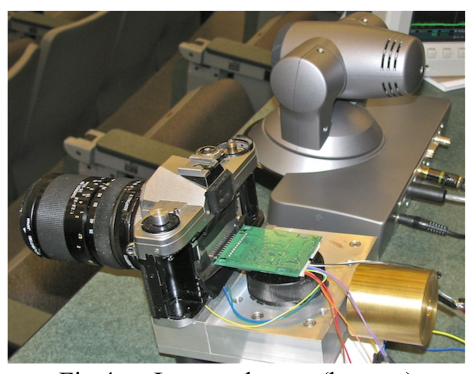

Fig 4a – Lens and array (bottom), PTZ video camera (top).

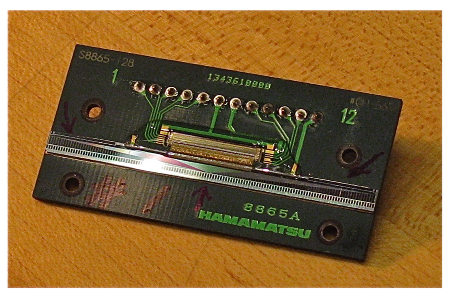

Fig 4b – Close up of array sensor.

Linear Photodiode Array Sensor and PTZ Camera:
The linear array of photodiodes consists of 128 horizontal photodiodes and has physical dimensions of 5 x 2.5 cm (Fig 4b). The IR light from the LEDs strike the photodiodes in the film plane of a still camera via a converging lens (Fig 4a). In our testing the scan refresh rate was 448Hz, but this particular array can go up to 4kHz [2].

ACCURACY

To test the system's accuracy, a single LED (Fig 5b) was affixed to a train set 22 feet from the linear array sensor. The target's position was measured and compared to the linear array's calculated position at 0.5 inch intervals. Fig 5a plots actual horizontal position versus calculated horizontal position. In the worst case, the same calculated position is seen for four consecutive locations of the LED necklace. Each data point is separated by 0.5 inches, therefore **the system's accuracy is ±2 inches at 22 feet**. Fig 5c uses the pinhole approximation to show the optical principle behind the linear array sensor.

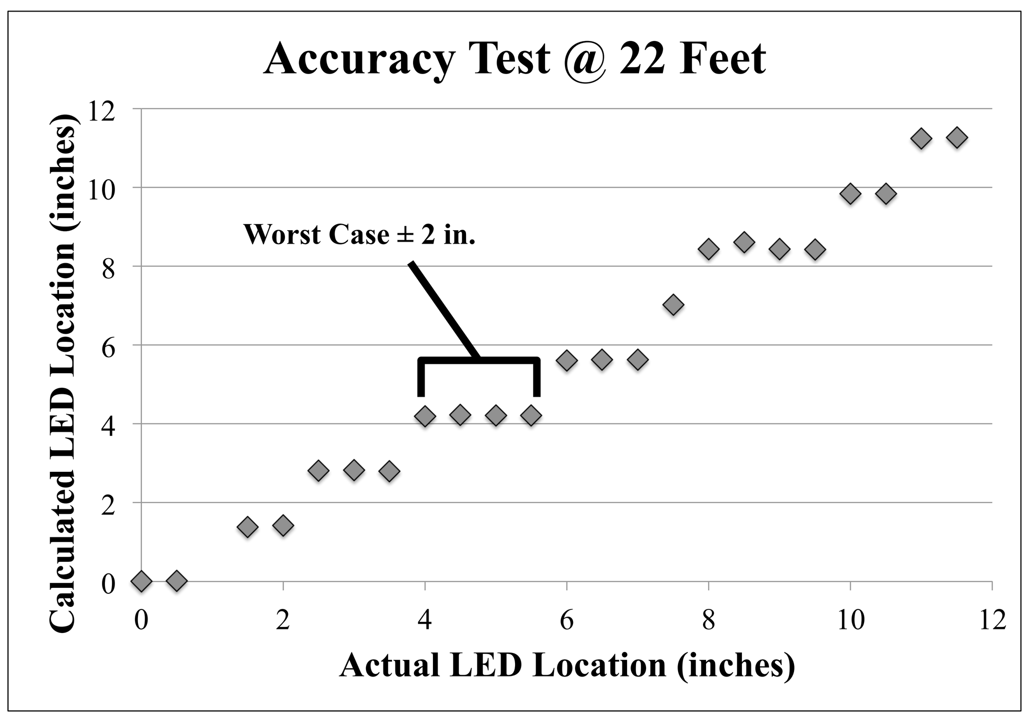

Fig 5a – Accuracy graph for a single LED.

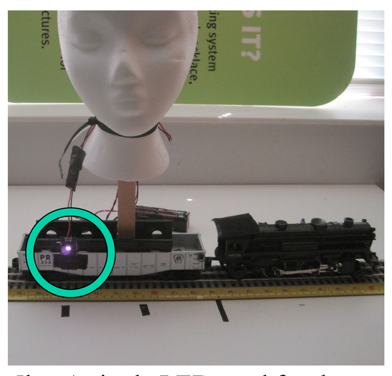

Fig 5b – A single LED used for the accuracy test.

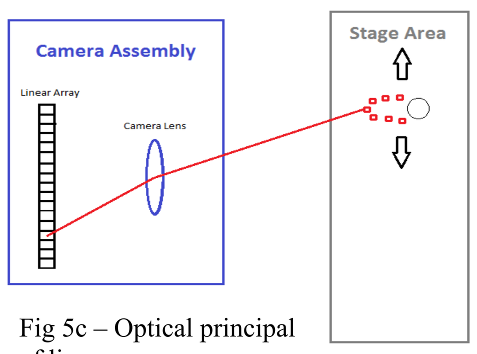

Fig 5c – Optical principal of linear array sensor.

NOISE FILTERING

A presenter tracking system that tracks an IR necklace must be able to filter noise IR signals to be considered a robust system [3]. The method used to filter the non-necklace IR signals is a digital band-pass filter. The output voltages from each of the 128 photodiodes are recorded for 4-5 cycles of the LEDs' flashing, and then put through the band-pass filter. Then, if any of the 128 diodes signals pass a predefined threshold, a panning command corresponding to the speaker's location is sent to the recording camera.

Fig 6a – Partially zoomed in image of large lecture hall.

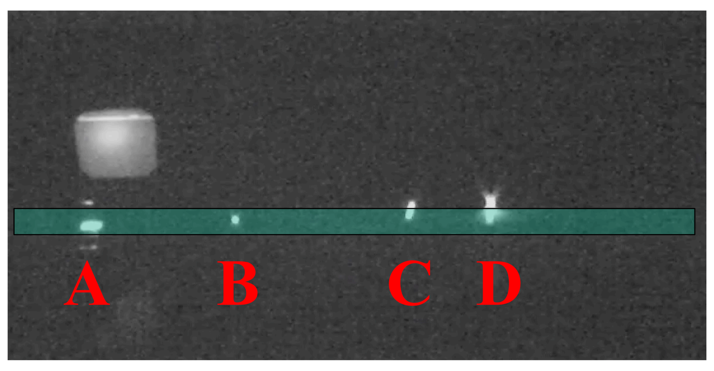

Fig 6b – Wide-angle IR image of stage. The green bar represents range seen by array.

Fig 6c – Graphical representation photodiodes seeing IR light. The top graph is before the band-pass filter, the bottom is after.

The filtering capability of the linear array system was tested in the large lecture hall shown in Fig 6a. The array was 44 feet from the speaker at center stage. Fig 6b is an IR image of the entire stage with the following objects labeled as: A–projector, B–flashing desklamp (4Hz), C–LED necklace (70Hz), D–DC desklamp. The graphs in Fig 6c plot diode number [0-127] (x-axis) versus voltage (y-axis). The top graph is unfiltered, and all the IR sources are shown. The bottom graph is filtered and shows only the 70Hz necklace.

VERTICAL POSITION TRACKING

The principal drawback of this system is its inability to track the vertical position of the IR LEDs. A single LED was tested to have an allowable **vertical movement range of 3.5 inches at a distance of 22 feet** from the linear array sensor. In most cases it is assumed that the height of the speaker will not change, and the system performs well. However, rooms with elevated stages, angled stages, or steep stadium seating present difficulty when the speaker travels the depth of the stage.

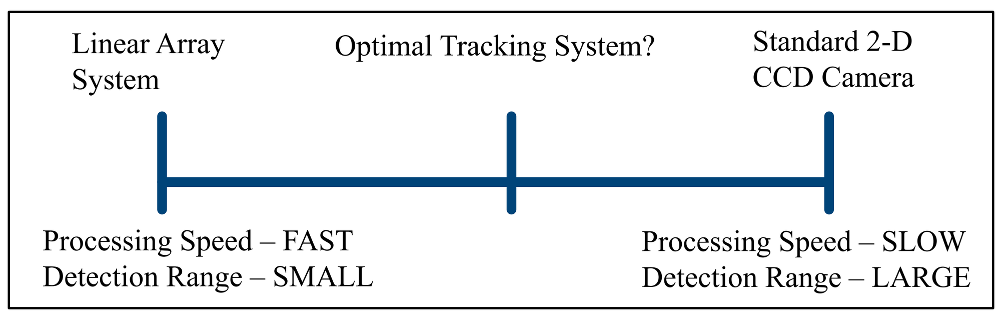

Fig 7 – Diagram showing the inverse relationship between detection range and processing speed.

Fig 7 suggests an optimal medium between a linear array system and a video camera with a full 2-D CCD sensor. The reason that a full 2-D CCD sensor is not used with this technique is that it is too processing intensive (128 vs. >100,000 pixel values). A more robust system may **stack 3-10 linear arrays on top of one another** in the film plane to allow for greater change in vertical positioning of the speaker, which frequently happens in certain room environments.

SYSTEM IN ACTION

A demonstration video of the linear array tracking and recording system is available from our research webpage: 'http://atlascollab.umich.edu/' under Projects → Robotic Tracking. This demo was done in the same lecture hall seen in Fig 6a, and shows that the system does a good job of following the speaker as he walks back and forth along the stage.

Bibliography:
[1] The ATLAS Experiment: http://www.atlas.ch/
[2] Hamamatsu S8865-128: http://sales.hamamatsu.com/en/products/solid-state-division/si-photodiode-series/si-photodiode-array/part-s8865-128.php
[3] "University of Michigan lecture archiving and related activities of the U-M ATLAS Collaboratory Project". J Herr *et al, 2008, J. Phys.: Conf. Ser.* **119** 082005