Lecture Adaptation for Students with Visual Disabilities Using High-Resolution Photography

Gregory Hughes
gregory.hughes@cl.cam.ac.uk

Peter Robinson
pr@cl.cam.ac.uk

Computer Laboratory, University of Cambridge
15 JJ Thomson Avenue, Cambridge UK CB3 0FD

ABSTRACT

Visual content in lectures can be enhanced for use by students with visual disabilities by using high-resolution digital still cameras. This paper presents a system which uses two high-resolution cameras; one to capture multiple sources of visual content and another to monitor the head pose of up to 20 audience members. This capture technique eliminates the need for multiple cameras or intrusive and distracting instrumentation but introduced some new problems which were solved with an algorithm used to distinguish between two possible sources of visual interest.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Object recognition, Tracking*; K.3.1 [**Computers and Education**]: Computer Uses in Education—*Computer-assisted instruction*

General Terms

Human Factors

Keywords

students with disabilities, visual disabilities, time-lapse photography

1. INTRODUCTION

Students with visual impairments struggle in a classroom or lecture setting due to the inherent visual nature of information presented on a whiteboard, blackboard, or a digital display. Students generally sit between 4 and 20 metres away from the visual information, sometimes making it difficult even for an average-sighted student to read lecturers' writing, and impossible for students with visual impairments. Many American universities provide note-takers for students with disabilities who supply the students with a copy of the notes after the lecture. This approach does not provide the student with access to the visual information during the lecture. Previous results have shown that visual information can be captured and enhanced in real time using high-resolution digital cameras[2]. This paper presents a system which is able to capture and display visual information as well as determine between two possible sources of visual information by tracking the head pose of audience members. Head-pose estimation is accomplished in an unobtrusive manner and requires no per-user calibration.

ASSETS'06, October 22–25, 2006, Portland, Oregon, USA.
ACM 1-59593-290-9/06/0010.

2. PREVIOUS WORK

Our system employs computer vision techniques and time-lapse photography to enhance the visual information presented in a lecture for use by a student with a visual disability in real time. High-resolution time-lapse photography is used instead of video because it allows more information to be captured with a single camera. Zooming is trivially accomplished without blurring due to the high-resolution images of the visual information. Areas of interest, such as a whiteboard or projector screen, are marked by the user. Given the four corners of any area, a perspective transformation is computed to compensate for the angle between the visual information and the camera. If the source is a whiteboard or overhead projector (OHP) then the next step is to enhance the material by running an adaptive threshold algorithm developed by Wellner[6]. Obstructions can then be removed and new text or additions are highlighted. The result of this process provides a very legible copy of the visual information which can be zoomed and changed to suit the needs of any individual with a visual impairment (See Figure 1).

Many lecture theatres have multiple sources of visual information such as a digital projector and an OHP. Thus it is necessary to develop a method of indicating the most relevant source of visual information at any given point in time. Stiefelhagen *et al* designed a system to record and index meetings by using head pose information to determine which attendee was speaking[4]. Their system, like many commercially available systems, requires per-user calibration which would not be achievable in a standard lecture environment.

3. DETERMINING VISUAL POINT OF FOCUS

To determine the visual point of focus in a lecture theatre we developed a head-pose estimation algorithm which is able to compute the aggregate head pose of an audience without requiring per-user calibration. An 8 Megapixel (MP) Canon S80 is used to capture images of the audience every three seconds. Faces are then extracted from the image using the Robust Real-Time Face Detection[5] algorithm. Once faces are isolated it is necessary to detect facial features which can then be used to compute an estimated head pose.

In our algorithm we chose to use the four corners of a

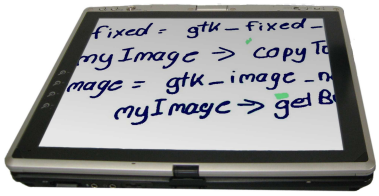

Figure 1: The user interface provides a scrollable high-resolution image of the visual content presented in a lecture.

(a) (b) (c) (d) (e)

Figure 2: Shows the process by which the corners of a subject's eyes are located.

subject's eyes to determine head pose. Finding the corners of a subject's eyes was done by first computing a high and low resolution symmetry map using a fast radial symmetry detection algorithm[3]. The low resolution symmetry map (Figure 2(a)) provides a rough estimate as to the location of a subject's eyes. The high-resolution symmetry map (Figure 2(b)) will show details such as the pupils and the corners of the eyes. These two maps are then combined and overlapping areas are marked in the top two quadrants of the image (Figure 2(c)). Eyebrows are ignored by removing the top third of the overlapping regions. The remaining overlapping portions correspond directly to the subject's two eyes, and thus their outer edges can be marked as they correspond to the corners of the eyes (Figures 2(d)–(e)).

Given the four corners of an individual's eyes, it is possible to recover head yaw (i.e. left to right movement) because of two basic assumptions[1]:

- Two eyes on any individual are the same size.
- The four corners of an individual's eyes are collinear.

With this information, we can determine whether an individual is looking left or right by comparing the distance between the two corners of the left eye to the distance between the two corners of the right eye.

3.1 Experiment

Two tests were conducted in a lecture theatre within the William Gates Building at the University of Cambridge. A high-resolution camera was connected to a computer which took one picture every three seconds. In each test, six mock students who were instructed to look at either screen one or screen two for varying lengths of time. The subjects were instructed either verbally or with the use of a laser pointer.

3.2 Results

For each face, the system would return a result of left, right, or indeterminable. During the two tests 79 out of 937 detected faces were identified incorrectly, resulting in an error-rate of about 8%. If we consider the faces marked indeterminable as false results, then it increases this error-rate to about 10%. Therefore, while tracking audience members, the system was able to distinguish between left and right focal attention with approximately 90% accuracy on an individual basis. Using these results, the binomial probability formula predicts that our algorithm will correctly identify the head pose of a majority of audience members in an audience of 15 detected faces with 99.997% accuracy. Thus, a good assumption can be made as to the point of visual focus in a lecture theatre.

The frequency at which images were captured presented a problem in real-world tests as three second intervals seemed to be too infrequent to produce an accurate representation of visual focus in real time. Decreasing the interval can easily be accomplished by the use of a high-resolution video camera, such as the 8 MP AVT Oscar, and parallel computing. The system also lacks the ability to distinguish between the lecturer and the screen which is behind him or her. This limitation can only be overcome by increasing the accuracy of the gaze estimation to approximately 2° allowing the intersection of gaze vectors to be accurately computed. To obtain this degree of accuracy the system would have to track the pupils and head movements, requiring a more intrusive tracking method. As the system is designed to distinguish between two screens, it is equally beneficial to know which screen the audience is looking at as it is to know which screen the audience is not looking at.

4. CONCLUSION

We were able to develop a system composed of two primary parts, one which enhances the visual information and the other which accurately determines where an audience is looking given two possible sources of visual attention. The system is reasonably inexpensive as it only requires off-the-shelf PC hardware and consumer-level digital cameras. Further tests will be required to see if the response of the system matches where a lecturer thinks the audience should be obtaining their visual information. Other tests will be needed to determine the usefulness of this system to a student with a visual disability obtaining enhanced visual information.

5. REFERENCES

[1] T. Horprasert, Y. Yacoob, and L. S. Davis. Computing 3-d head orientation from a monocular image sequence. In *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG '96)*, page 242, Washington, DC, USA, 1996. IEEE Computer Society.

[2] G. Hughes and P. Robinson. Time-lapse photography as an assistive tool. In *Proceedings of the 3rd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT)*, pages 87–89, Cambridge, UK, 2006.

[3] G. Loy and A. Zelinsky. A fast radial symmetry transform for detecting points of interest. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, pages 358–368, London, UK, 2002. Springer-Verlag.

[4] R. Stiefelhagen, J. Yang, and A. Waibel. Modeling focus of attention for meeting indexing. In *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, pages 3–10, New York, NY, USA, 1999. ACM Press.

[5] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, 2004.

[6] P. D. Wellner. Interacting with paper on the DigitalDesk. Technical Report UCAM-CL-TR-330, University of Cambridge, Computer Laboratory, Mar. 1994.