Back to Paper: A Technique for Browsing Multimedia Information by Pointing on Handwritten Notes

Ozan Cakmakci[1] and François Bérard[2]

Abstract

We propose a system for a lecture room with a camera recording the lecturer and an overhead camera observing the locations of a student's writing. Based on knowing the time and location of the handwritten notes on ordinary paper, we are able to create a novel method for capturing and accessing the lecture content in real-time (this doesn't require any pattern or stroke recognition). Essentially, our system provides the ability to browse in time the recording of a presentation using the notes taken during that presentation. We strive to leverage off of paper as a familiar medium and provide services related to augmenting the user's note taking and reviewing experience.

1. Introduction and Related Work

Paper has several desirable attributes that justify its widespread use. Paper is relatively cheap, requires no electrical power to display its contents, makes clever use of ambient light and has a very wide viewing angle while maintaining an acceptable contrast ratio. Alongside these technical arguments, K12 education still emphasizes the use of paper as the primary medium for note taking in class, deeply rooting the extensive use of paper in the society. Due to these technical, social, and economic reasons, we believe that ordinary paper has a place in ubiquitous computing, at least within the foreseeable future. Similar systems have been developed, some of them require vendor specific paper (Anoto pen) or a digitizing tablet (Audio notebook) and others are focused on off-line processing of annotations or documents ([2] or [4]) and none of them focus on the real-time nature of the capture and access process with minimal hardware requirements (overhead camera and an accelerometer based pen). In our system, the color coding of the paper is used primarily to establish a coordinate system for tracking the pen and not to determine the 3d position and orientation of the markers as done in [3].

2. Capturing and Accessing Specific Moments in a Meeting

Creating useful indices in order to facilitate the access of captured multimedia material is a well known problem in ubiquitous computing [1]. In this paper, we do not address the automatic capture portion of the proposed interface. A useful index in our application is knowing *when* people write (preferably each word) on paper, along with *where* they write at that moment. For the specific set of French lecture notes shown in Fig. 1, we manually created 3 regions: "salutations", "nombres" and "formes". Each region consisted of a rectangle in the paper coordinate system. When the student goes home to review that lecture, by pointing the marked pen

[1] College of Optics and Photonics: CREOL & FPCE, University of Central Florida, Orlando, Florida, USA
[2] Human Computer Interaction Groups (IIHM), University of Joseph Fourier, Grenoble, France

over any of these indexed regions would rewind or forward the lecture video to the start of the appropriate section and start playing the lecture video. The ambiguity between pointing and navigation is resolved through a time based filter where the system waits a certain amount of time before rewinding/forwarding the video to the appropriate lecture time. The following section presents techniques to track the tip of the pen with respect to a coordinate system on a piece of paper. Our system provides the ability to browse in time the recording of a presentation using the notes taken during that presentation.

3. System Implementation

We are applying well-known techniques from computer vision to develop a novel interaction technique by finding out where on paper the user writes or is pointing to using a marked pen or pencil. A ceiling mounted camera captures images of the workspace of the user, see Fig. 1 as an example input image to our system. For a description of the complete computer vision algorithm, at the block diagram level, see Fig. 2.

Fig. 1. Camera view looking at pen and paper (left), segmented green fiducials (right)

For each frame, the algorithm applies the statistical color models for segmentation, connected components on the segmented image, finds correspondences between the camera and the real world points, estimates the homography matrix based on the correspondences and projects the centroid of the pen onto the paper plane in order to estimate the position of the pen with respect to the coordinate system on the paper plane.

In order to locate the paper, we print six green fiducials on the paper using a color printer. We build statistical color models for the green fiducials on paper and red fiducial on the pen. Normalized RGB is used as the color space which means diving each individual pixel by the luminance (R+G+B).

$$p(\text{color}|\text{pixel}) = \frac{p(pixel \mid color)\, p(color)}{p(pixel)} \quad (1)$$

Using bayes rule, we define the segmentation problem as in equation (1).

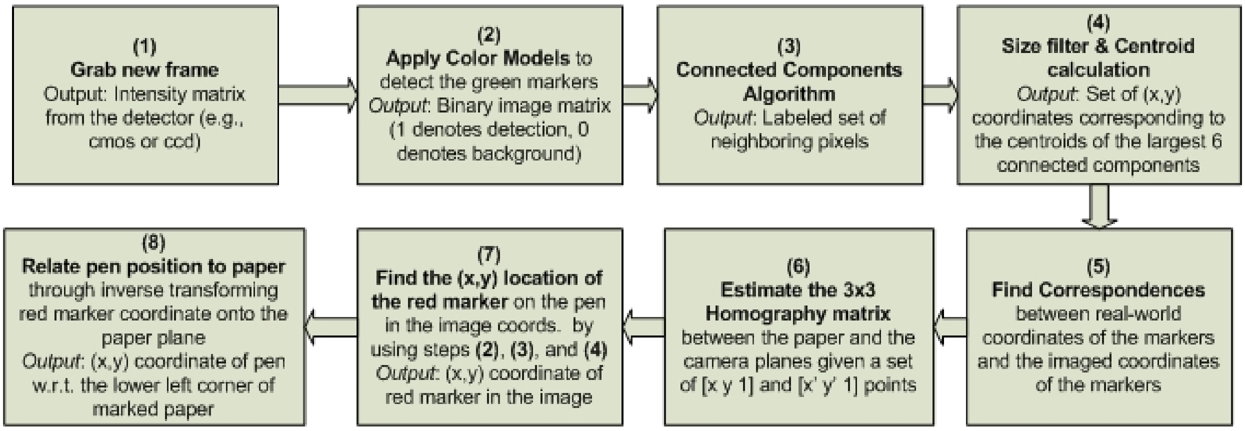

Fig. 2. Block diagram of a complete cycle through the vision algorithm.

When the system is running, it segments the green fiducials based on probabilistic color models, applies connected component algorithm and calculates the area of the largest six connected components. Based on the segmentation, we use a distance based sorting criteria to solve the correspondence problem between the points in the image plane and on paper. After obtaining points of correspondence we estimate the homography matrix between the image plane and the paper plane, which we will briefly describe. The perspective transformation between the paper plane and the image plane can be described as follows:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \qquad (2)$$

[x y 1]^T represents the vector of world points and **[x' y' 1]**^{T} represents the vector of image points corresponding to those world points. H represents the 3x3 transformation matrix. Each correspondence between the image coordinate system and the world coordinates will give us 2 equations. We have 8 unknowns, therefore, we will need at least 4 correspondences to estimate the parameters of the homography matrix. We can setup the following matrix using the correspondences:

$$
\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1'\cdot x_1 & -x_1'\cdot y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y_1'\cdot x_1 & y_1'\cdot y_1 \\ & & & & \vdots & & & \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \\ x_3' \\ y_3' \\ x_4' \\ y_4' \end{bmatrix} \quad (3)
$$

Since equation 2 is of form Ax=b, there exist several techniques to solve it. One could use the pseudo-inverse or minimize the residuals. After estimating the homography matrix, we locate the marked writing utensil via it's color. Once we segment the largest blob representing the tip of the pen, we take the centroid of this region and inverse transform that point onto the paper plane, using the estimated homography matrix, in order to find the position of the pen with respect to the paper coordinate system.

4. Conclusions and Future Work

We developed an interface where a pen is utilized spatially to access previously indexed multimedia documents in a meeting. The system is based on ordinary paper and pen with color markers. We have designed and implemented the access part of the interface via standard techniques from computer vision. The indexing has been done manually so far in our experiments. We are planning on equipping a pen with an accelerometer to investigate whether we can create word level timestamps based on accelerometer data during the capture phase. At this phase, the system does not support multiple users or multiple pages of paper. Automatic indexing will be the major extension in the future.

5. References

[1] Gregory D. Abowd and Elizabeth D. Mynatt. Charting Past, Present and Future Research in Ubiquitous Computing. *ACM Transactions on Computer-Human Interaction*, Special issue on HCI in the new Millenium, **7**(1):29-58, March.

[2] Heiner, J.M., S.E. Hudson, and K. Tanaka. Linking and messaging from real paper in the Paper PDA.Proceedings of UIST'99, pp. 179 - 186.

[3] J. Rekimoto and Y. Ayatsuka. CyberCode: Designing Augmented Reality Environments with Visual Tags. Proc. DARE 2000 on Designing augmented reality environments, pp. 1-10.

[4] J. Kim, S. Seitz and M. Agrawala. Video-Based Document Tracking: Unifying your Physical and Electronic Desktops. Proc. UIST 2004, pp.99-107.