VISUAL ECHO CANCELLATION IN A PROJECTOR-CAMERA-WHITEBOARD SYSTEM

Hanning Zhou† *Zhengyou Zhang*‡ *Thomas Huang*†

† University of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801
‡ Microsoft Research, One Microsoft Way, Redmond, WA 98052
† {hzhou, huang}@ifp.uiuc.edu, ‡ zhang@microsoft.com

ABSTRACT

We propose to incorporate a whiteboard into a projector-camera system. The whiteboard serves as the writing surface (input) as well as the projecting surface (output). The ability to write and draw on top of computer-projected content opens up many new opportunities for real-time collaborations between people located on-site and remotely. Such applications inevitably require extracting handwritings from video images that contain both handwritings and the projected content. By analogy with echo cancellation in audio conferencing, we call this problem *visual echo cancellation*. This paper presents one approach to accomplish the task. Our visual echo cancellation algorithm estimates the incident light and derive the surface albedo based on both incident light and refection. By estimating the albedo, we can extract the writings and recover their colors. Our approach includes two basic components of projector-camera systems: geometric calibration and color calibration. The first one solves the mapping between the position in the camera view and the position in the projector screen, while the second one solves the mapping between the actual color of the projected content and that seen by the camera.

1. INTRODUCTION

During the past few years we witnessed the transformation of video cameras and projectors from expensive lab equipments to affordable consumer products. This triggers many human-computer interaction systems that incorporate both the large-scale display provided by the projector and intelligent feedback from one or more cameras [1, 2, 3, 4, 5]. On the other hand, the whiteboard is still an indispensable part of many meetings (including lecturing, presentation and brainstorming), because it provides a large shared space for the participants to focus their attention and exchange their ideas spontaneously [6, 7]. One can write or draw his/her idea on it with an easily accessible marker. Therefore we propose to integrate the whiteboard into a projector-camera system by using it as both the writing surface and the projecting surface. Several immediate advantages are:

1. Computer presentations (such as PowerPoint) and whiteboard discussions are seamlessly integrated into one session. Meeting attendees will not be distracted by switching from the screen to the whiteboard, and vice versa.

2. Such a system enables local and remote attendees to collaborate with each other on a shared workspace. Local attendees have a much more natural writing surface than most commercial large display products.

3. Most importantly, the system can be easily deployed on top of current meeting environments. It is therefore much more economical than most large display products that requires installing expensive equipments and accessories.

The idea of integrating a writing surface into projector-camera systems dates back to 1993 [8, 9, 10, 11, 12, 13, 14]. Most of them targeted at remote sketching in the office desk scenarios while our proposed application is more universal in that the user can use the whiteboard as a writing/drawing/anotating desktop as well as a presentation screen for both on-site and remote collaborations.

In our case, since the captured video contains both writings on the physical whiteboard and contents projected from the computer, it is very important to separate whiteboard writings from the projected contents. Some of the benefits are:

1. It dramatically reduces the bandwidth requirement for teleconferencing, because both extracted writing and the computer-projected contents can be transmitted with very low bandwidth, comparing with the original mixed video, since the video is affected by shadow and lighting variation.

2. It considerably improves the remote users' experience in teleconferencing in several ways, to be discussed below.

This work was done when Hanning Zhou was an intern at Microsoft Research.

3. Extracted writings are essential for archiving and browsing meetings of Writing on the whiteboard usually indicates an important event in a meeting.

4. By feeding the results to an OCR (Optical Character Recognition) system, the meeting archive can be more easily accessed and transferred into other forms.

By analogy with echo cancellation in audio conferencing, we call this problem visual echo cancellation. This problem is related to, but more complicated than the task of extracting writing from a clean background that appears in previous works [12, 15].

Visual Echo, by strict definition, is the appearance of the projected contents viewed by the camera. Visual Echo Cancellation is defined as extracting the physical writings from the video containing both the writings and the visual echoes. In order to achieve this goal, we need an accurate prediction of the appearance of the computer projected content as viewed by the camera. This requires two basic components:

1. Geometric calibration: It concerns the mapping between the position in the camera view and the position in the projector screen.

2. Color calibration: It concerns the mapping between the actual color of the projected content and that seen by the camera.

For geometric calibration, we assume that both camera and projector are linear projective, and implement a robust, accurate and simple technique by leveraging the fact that the projector can actively project the patterns we want. For color calibration, we model pixels on the visual echo as independent Gaussian random variables and propose a lookup-table-based approach. Note that both components are useful for other projector-camera systems.

Researches on the basic techniques of geometric calibration date back to 1986 [16, 17]. They have been widely applied in projector-camera systems, ranging from virtual environment scenarios [18, 5] to presentation scenarios [2, 3].

Although color calibration has been studied in the context of achieving photometric uniformity, our application requires most accurate chromatic estimation between the original system color and the visual echo, while previous methods mainly attack intensity variations among multiple projectors.

The remainder of the paper is organized as follows. Section 2 gives an overview of our projector-camera-whiteboard system. Section 3 describes the geometric calibration. Section 4 introduces our color calibration method. Section 5 describes our visual echo cancellation method. Section 6 provides experimental results in both quantitative and visual forms. Section 7 summaries our contributions and limitations of the system.

2. A PROJECTOR-CAMERA-WHITEBOARD SYSTEM

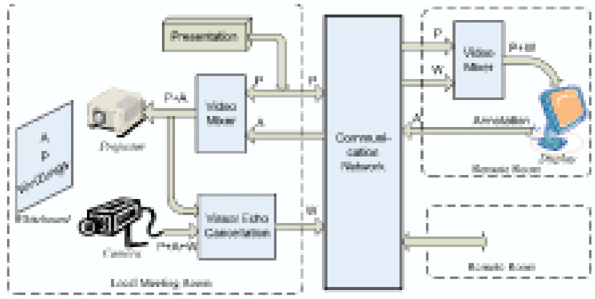

Fig. 1. A projector-camera-whiteboard system

Figure 1 illustrates how our projector-camera-whiteboard system works. The local meeting room is equipped with a projector, a camera, and a whiteboard. The projector and the camera are rigidly attached to each other, although theoretically they can be positioned anywhere as long as the projector projects on the whiteboard and the camera sees the whole projection area. The projector and the camera are linked to a computer, and the computer is connected to the communication network. Remote attendees also connect their computers to the communication network.

A presentation could be PowerPoint slides, a spreadsheet, a PDF file, etc. The data stream for the presentation is denoted by "P" in the figure. Remote attendees may annotate the presentation, and the annotation stream is denoted by "A". Both "P" and "A" are mixed together before sending to the projector for projecting on the whiteboard. During the presentation, the presenter or other local attendees may write or draw on the whiteboard. The camera captures both the projected content and the writings. Through geometric and color calibrations, the system predicts the appearance of the projected "P" and "A" viewed by the camera, i.e., the visual echo. The Visual Echo Cancellation module tries to extract only the writings on the whiteboard, denoted by "W", by subtracting the predicted visual echo from the live video. At the remote side, the presentation stream "P" and the whiteboard writing stream "W" are mixed before displaying on the computer.

3. GEOMETRIC CALIBRATION

Assuming that both camera and projector are linear projective and that the whiteboard surface is planar, it can be easily shown that the mapping between a point in the camera view and a point in the projector screen is a homography, and can be described by a 3×3 matrix H defined up to a scale factor. The idea of geometric calibration is to leverage the fact that the projector can actively project the patterns we want. The whole process takes less than 2 minutes and is only necessary when camera is moved with respect to the projector. The main steps are: 1. Sequentially project N ($N = 40$ in our implementation) rectangles and simultaneously capture their images using a fixed camera. 2. Detect the 4 corners of each rectangle in the images. 3. Use the $4 \times N$ detected corners and their corresponding known positions in the projector space to estimate the homography between the projector screen and the image plane of the camera.

4. COLOR CALIBRATION

For visual echo cancellation, for a given pixel in the projector space, we know its corresponding position in the camera space through geometric calibration described above; furthermore, we need to know what the corresponding color should look like in the captured video, and this is the task of color calibration. Note that the same color in the projector space appears different in the camera, depending where the color is projected on the whiteboard. This is because the projector lamp does not produce uniform lights, the lighting in the room is differing and not uniform, and the whiteboard surface is not Lambertian. Therefore, color calibration should be both color- and position-dependent.

Below are the main steps for color calibration: 1. Quantize the RGB color space into $9 \times 9 \times 9 = 729$ bins. 2. Project each quantized color over the whole display region and capture its image in synchronization. We store n ($n = 5$) frames for each color. 3. Rectify using the geometric calibration and divide the display region evenly into $32 \times 32 = 1024$ rectangular blocks. 4. Calculate the mean and variance of each color in each block across the n frames.

In this way, we build a lookup table for the 729 quantized colors at each of the 1024 blocks. Note that the spatial dimensionality is necessary because the same projected color will have different appearance at different position, as the second row in Figure 3 shows.

Given an arbitrary display content, we estimate the visual echo E by: first substituting each pixel with its corresponding mean color in the lookup table[1], and then backward-warping it to the camera view. To estimate the error bound for each pixel, we also lookup and warp the variance of the incident illumination of each pixel to get a pixel-wise variance map V.

5. VISUAL ECHO CANCELLATION

Figure 2 shows the flowchart of the visual echo cancellation process.

By writing/drawing with a paint marker on the whiteboard, we actually change the surface albedo of the whiteboard, and subsequently change the re Therefore extracting the writings boils down to detecting the changes of the surface albedo.

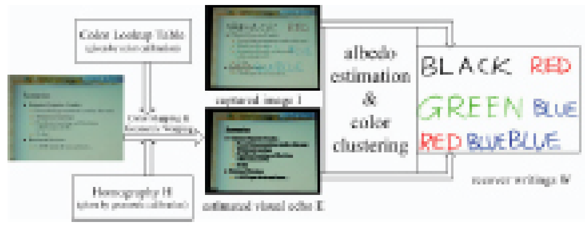

Fig. 2. Flowchart for visual echo cancellation

Assuming all the images are geometrically aligned, and denoting the incident light map by P, the surface albedo of the whiteboard by A, the pixel-wise color transformation due the camera sensor by C, and the visual echo by E, we have $E = C \times A \times P$. If nothing is written on the whiteboard, then the captured image I should be equal to E. If there is any thing written on the whiteboard, the surface albedo changes, and is denoted by $\tilde{A}$. The captured image can then be described by $I = C \times \tilde{A} \times P$. We can compute the albedo change by estimating the albedo ratio $a = \tilde{A}/A$ of the pixel $[x; y]$ in color channel $c \in \{R; G; B\}$, which is given by $a_{[x;y];c} = \frac{I_{[x;y];c}}{E_{[x;y];c}}$. Note that writings on the whiteboard absorb the lights, so $\tilde{A} \le A$, and in consequence $a_{[x;y];c} \le 1$.

Based on the albedo a, we can detect the writings and recover their colors. The albedo for the whiteboard region without writings should be 1. Assuming the sensor noise on the albedo is additive and has a zero-mean Gaussian distribution with variance $\frac{V}{E}$, we choose the following decision rule among various options based on experimental evaluations:

Pixel $[x; y]$ belongs to the written region if and only if

$$1 - \frac{a_{[x;y];R} + a_{[x;y];G} + a_{[x;y];B}}{3} > \frac{V_{[x;y];R} + V_{[x;y];G} + V_{[x;y];B}}{E_{[x;y];R} + E_{[x;y];G} + E_{[x;y];B}} \quad (1)$$

Note that the decision rule is one-sided, because, as mentioned earlier, the albedo ratio for written whiteboard region is less than or equal to 1.

For each pixel $[x; y]$ belonging to the written region, we can recover the writings with their colors as $W_{[x;y];c} = a_{[x;y];c} \times 255$, assuming the color intensity ranges from 0 to 255. Practically, due to the noise in geometric calibration, I and E are not exactly aligned. The 1 to 2 pixel errors are most evident near strong edges in E. Therefore in written region segmentation, we first apply an erosion on E, which increases the dark region. Thus the pixels near the dark regions in E have higher A and are less likely be classified as written region. This preprocessing reduces error because in order to make their writings more visible, most users pre-

[1] For colors not in the table, we use linear interpolation of the two nearest bins.

fer to write on top of brighter background instead of darker background.

6. EXPERIMENTAL RESULTS

We tested our geometric calibration method using various projectors and video cameras, under both arti cial lighting and natural lighting conditions. The tting error for solving the homography based on correspondences ranges from 0.3 to 0.7 pixels.

For color calibration, we use a SONY projector and EVI30 camera. Comparing the estimated visual echo E with the actual captured image I, the average error is around 3 (the color intensity range is 0 » 255). The majority of the discrepancy is around the regions with strong edges, due to the noise in geometric calibration.

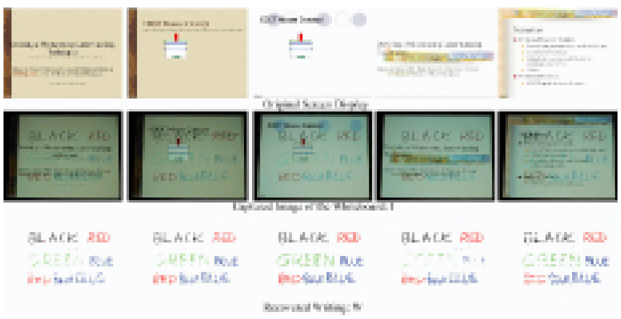

Fig. 3. Experimental results for Visual Echo Cancellation

Figure 3 shows the visual echo cancellation results on various backgrounds. One can see that majority of the writings are recovered. The only exceptions are the parts on top of the extreme complex background contents like texts in the presentation slides. However, the user rarely write over texts.

7. CONCLUSIONS

We have de ned the problem of visual echo cancellation in projector-camera-whiteboard systems and proposed a solution using both geometric calibration and color calibration. Visual echo cancellation has wide applications in real-time collaboration tasks, both on-site and remotely. The algorithm is tested on various backgrounds and display contents, and good results are achieved.

Both geometric and color calibrations could be used for other purposes. The geometric calibration technique has actually been integrated into our projector-camera based human-computer-interaction system, which tracks the image position of the laser dot to command the mouse cursor on the display screen.

For color calibration, projecting and capturing 729 £ n (= 3645 when n = 5) frames at 10 fps (to ensure projecting and capturing are synchronized) takes about 6 minutes. Fortunately, it is an off-line process, and only needs to be done once.

8. REFERENCES

[1] R. Surati, Scalable Self-Calibrating Display Technology for Seamless Large-Scale Displays, Ph.D. thesis, Massachusetts Institute of Technology, 1999.

[2] Rahul Sukthankar, Robert G. Stockton, and Matthew D. Mullin, "Smarter presentations: Exploiting homography in camera-projector systems," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2001.

[3] Ramesh Raskar and Paul Beardsley, "A self-correcting projector," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hawaii, December 2001, pp. 504–508.

[4] Ramesh Raskar, Jeroen van Baar, Paul Beardsley, Thomas Willwacher, Srinivas Rao, and Clifton Forlines, "ilamps: Geometrically aware and self-con guring projectors," ACM Transactions on Graphics, vol. 22, pp. 809–818, July 2003, Also appears in ACM SIGGRAPH 2003 Conference Proceedings.

[5] Jeroen van Baar, Thomas Willwacher, Srinivas Rao, and Ramesh Raskar, "Seamless multi-projector display on curved screens," in Eurographics Workshop on Virtual Environments, May 2003, pp. 281–286.

[6] E. Saund, "Bringing the marks on a whiteboard to electronic life," in In Proc. Cooperative Buildings—Integrating Information, Organizations, and Architecture: Second Int. Workshop, CoBuild 99 (Lecture Notes in Computer Science 1670), N. Streitz, J. Siegel, V. Hartkopf, and S. Konimi, Eds. 1999, vol. V, pp. 776–779, Springer.

[7] Liwei He, Zicheng Liu, and Zhengyou Zhang, "Why take notes? use the whiteboard system," in In Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), Hong Kong, April 2003, vol. V, pp. 776–779.

[8] Pierre Wellner, "Interacting with paper on the DigitalDesk," Communications of the ACM, vol. 36, no. 7, pp. 86–97, 1993.

[9] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs, "The of ce of the future: A uni ed approach to image-based modeling and spatially immersive displays," Computer Graphics, vol. 32, no. Annual Conference Series, pp. 179–188, 1998.

[10] D. Hall, C. Le Gal, J. Martin, O. Chomat, T. Kapuscinski, and J. Crowley, "Magicboard: A contribution to an intelligent of ce environment," in Proc. of the International Symposium on Intelligent Robotic Systems, 1999.

[11] Elizabeth D. Mynatt, Takeo Igarashi, W. Keith Edwards, and Anthony LaMarca, "Flatland: New dimensions in of ce whiteboards," in CHI, 1999, pp. 346–353.

[12] Naoya Takao, Jianbo Shi, and Simon Baker, "Tele-graf ti: A camera-projector based remote sketching system with hand-based user interface and automatic session summarization," IJCV, 2003.

[13] J.L. Crowley, J. Coutaz, , and F. Berard, "Things that see," Communications of the ACM, pp. 54–64, July 2000, Also appears in ACM SIGGRAPH 2003 Conference Proceedings.

[14] F. Berard, "The magic table: Computer-vision based augmentation of a whiteboard for creative meetings.," in In Proc. International Workshop on Projector-Camera Systems (ProCams), Nice, France, October 2003, pp. In CD-Roms of the Ninth International Conference on Computer Vision (ICCV'03).

[15] Tegrity Inc., "The weblearner solution," .

[16] R.Y Tsai, "An ef cient and accurate camera calibration technique for 3d machine vision," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1986.

[17] Zhengyou Zhang, "A xible new technique for camera calibration," PAMI, vol. 22, no. 11, pp. 1330–1334, 2000.

[18] W. Krueger, C.-A. Bohn, B. Froehlich, H. Schueth, W. Strauss, and G. Wesche, "The responsive workbench: A virtual work environment," IEEE Computer, vol. 28, no. 7, pp. 42–48, July 1995.