NOTETAKING WITH A CAMERA:
WHITEBOARD SCANNING AND IMAGE ENHANCEMENT

Zhengyou Zhang and Li-wei He

Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
Email: {zhang,lhe}@microsoft.com

ABSTRACT

This paper describes a system for scanning the content on a whiteboard into the computer by a digital camera and also for enhancing the visual quality of whiteboard images. Because digital cameras are becoming accessible to average users, more and more people use digital cameras to take pictures of whiteboards instead of copying manually, thus significantly increasing the productivity. However, the images are usually taken from an angle, resulting in undesired perspective distortion. They also contain other distracting regions such as walls. We have developed a system that automatically locates the boundary of a whiteboard, crops out the whiteboard region, rectifies it into a rectangle, and corrects the color to make the whiteboard completely white. In case where a single image is not enough (e.g., large whiteboard and low-res camera), we have developed a robust feature-based technique to automatically stitch multiple overlapping images. We therefore reproduce the whiteboard content as a faithful electronic document which can be archived or shared with others. The system has been tested extensively, and very good results have been obtained.

1. INTRODUCTION

A whiteboard provides a large shared space for collaboration among knowledge workers. It is not only effective but also economical and easy to use – all you need is a flat board and several dry-ink pens. While whiteboards are frequently used, they are not perfect. The content on the whiteboard is hard to archive or share with others who are not present in the session. Imagine that you had a fruitful brainstorming session with all the nice drawings on the whiteboard, and you have to copy them in your laptop. If you have another meeting right after, you will not have time to copy the contents; if other people reserve the meeting room and use it right after, the contents on the whiteboard will be erased. Because digital cameras are becoming accessible to average users, more and more people use digital cameras to take images of whiteboards instead of copying manually, thus significantly increasing the productivity. The system we describe in this paper aims at reproducing the whiteboard content as a faithful, yet enhanced and easily manipulable, electronic document through the use of a digital (still or video) camera.

However, images are usually taken from an angle to avoid highlights created by flash, resulting in undesired perspective distortion. They also contain other distracting regions such as walls. The system we have developed uses a series of image processing algorithms. It automatically locates the boundary of a whiteboard as long as there is a reasonable contrast near the borders, crops out the whiteboard region, rectifies it to a rectangle with the estimated aspect ratio, and finally correct the colors to produce a crisp image.

Besides image enhancement, our system is also able to scan a large whiteboard by stitching multiple images automatically. Imagine that you only have a built-in camera with maximum resolution 640×480; this is usually not high enough to produce a readable image of a large whiteboard. Our usability study shows that we need about 25 pixels per inch[1] in order to read whiteboard images with normal writing. Our system provides a simple interface to take multiple images of the whiteboard with overlap and stitches them automatically to produce a high-res image. The stitched image can then be processed and enhanced as mentioned earlier.

The whiteboard scanning subsystem is similar to the ZombieBoard system developed at Xerox PARC [6]. The difference is that they reply on a pan-tilt video camera while we can use a free-moving (still or video) camera as long as there is an overlap between successive views.

In last year's ICASSP, we presented a whiteboard capture system for a conference room setup [4]. In that system, a high-resolution digital camera is mounted on the opposite wall of the whiteboard and fixed toward the whiteboard, and a microphone is installed in the middle of the table. Both whiteboard content and audio signals are captured during the meeting. The whiteboard image sequence is post-analyzed, and strokes and keyframes are produced and time-stamped. Therefore the whiteboard content serves as a visual index to efficiently browse the audio meeting. On the other hand, the system presented in this paper is very light-weight. It can be used to archive whiteboard content whenever the user feels necessary.

Please note that due to space limitation, the paper is written to present the high-level description of the system. The technical details, including original contributions such as whiteboard detection, whiteboard aspect ratio estimation, white balancing, hierarchical feature matching, and global image registration, are described in the accompanying technical report [8].

2. OVERVIEW OF THE SYSTEM

Before going further, let us look at Figure 1. On the top is an original image of a whiteboard taken by a digtal camera, and on the bottom is the final image produced automatically by our system. The content on the whiteboard gives a flow chart of our system.

As can be seen in Fig. 1b, the first thing we need to decide is whether it is enough to take a single image of the whiteboard. If the whiteboard is small (e.g., 40' by 40') and a high-res digital camera (e.g., 3 mega pixels) is used, then a single image is usually enough. Otherwise, we need to call the whiteboard scanning subsystem, to be described in Section 4, to produce a composite

---

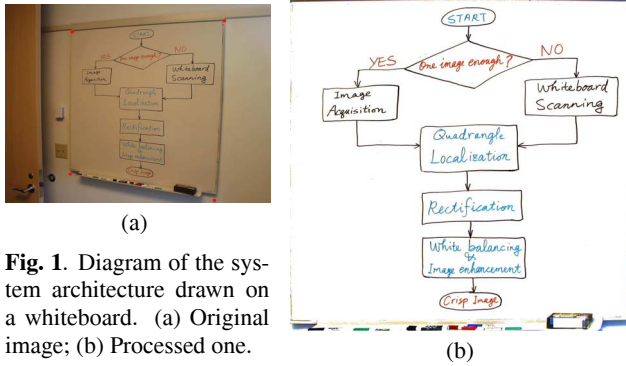[1]one inch (1') $\approx$ 2.54cm.

(a)

(b)

Fig. 1. Diagram of the system architecture drawn on a whiteboard. (a) Original image; (b) Processed one.

image that has enough resolution for comfortable reading of the whiteboard content. Below, we assume we have an image with enough resolution.

The first step is then to localize the borders of the whiteboard in the image. This is done by detecting four strong edges. The whiteboard in an image usually appears to be a general quadrangle, rather than a rectangle, because of camera's perspective projection. If a whiteboard does not have strong edges, an interface is provided for the user to specify the quadrangle manually.

The second step is image rectification. For that, we first estimate the actual aspect ratio of the whiteboard from the quadrangle in the image based on the fact that it is a projection of a rectangle in space. From the estimated aspect ratio, and by choosing the "largest" whiteboard pixel as the standard pixel in the final image, we can compute the desired resolution of the final image. A planar mapping (a 3×3 homography matrix) is then computed from the original image quadrangle to the final image rectangle, and the whiteboard image is rectified accordingly.

The last step is white balancing of the background color. This involves two procedures. The first is the estimation of the background color (the whiteboard color under the same lighting without anything written on it). This is not a trivial task because of complex lighting environment, whiteboard reflection and strokes written on the board. The second concerns the actual white balancing. We make the background uniformly white and increase color saturation of the pen strokes. The output is a crisp image ready to be integrated with any office document or to be sent to the meeting participants.

3. DETAILS OF WHITEBOARD IMAGE PROCESSING

We now provide details of the image processing techniques used in our system. The whiteboard scanning system will be described in the next section. Because of space limitation, full details will be provided in a technical report accompanying this paper [8].

3.1. Major Steps

Automatic Whiteboard Detection. As was mentioned in the introduction, this work was motivated by developing a useful tool to capture the whiteboard content with a digital camera rather copying the notes manually. If the user has to click on the corners of the whiteboard, we have not realized the full potential with digital technologies. The automatic technique of whiteboard detection is based on Hough transform [5], but needs a considerable amount of engineering because there are usually many lines which can form
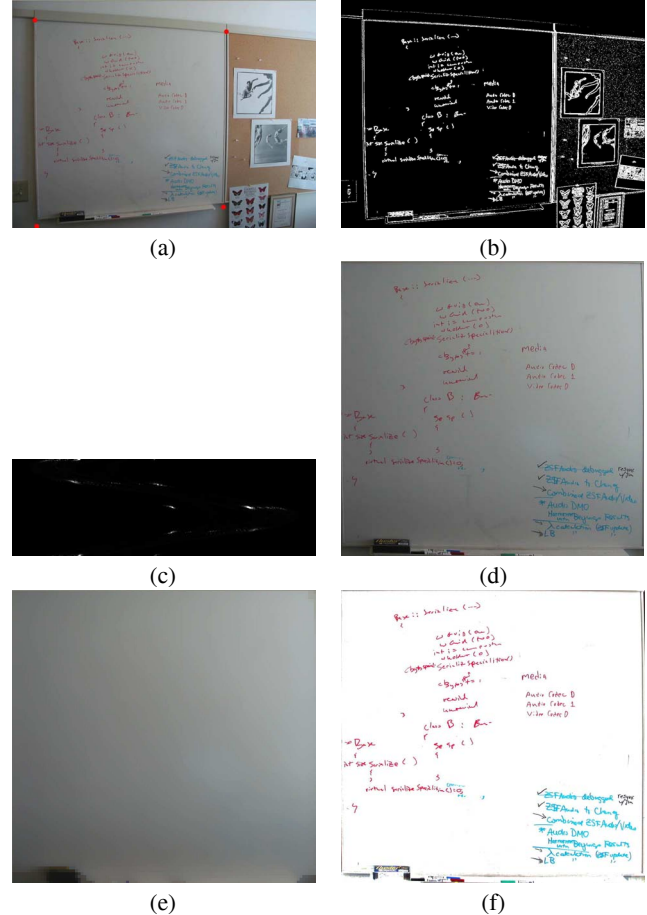

(a)

(b)

(c)

(d)

(e)

(f)

Fig. 2. Whiteboard image enhancement: (a) Original image together with the detected corners shown in small red dots; (b) Edge image; (c) Hough image with ρ in horizontal axis and θ in vertical axis; (d) Cropped and rectified whiteboard image; (e) Estimated whiteboard color; (f) Final enhanced image.

a quadrangle. Whiteboard corners are obtained by intersecting detected border lines. The red dots in Fig. 1a are the detected corners of the whiteboard. Note that the actual lower border of the whiteboard does not have strong edge information, so our technique reasonably detects the line corresponding to the pen holder.

Image Rectification. Because of the perspective distortion, the image of a rectangle appears to be a quadrangle. However, since we know that it is a rectangle in space, we are able to estimate both the camera's focal length and the rectangle's aspect ratio. The details will be given in our technical report [8]. The next task is to rectify the whiteboard image into a rectangular shape with the estimated aspect ratio. For that, we need to know the resolution of the final image. We determine the size in order to preserve in the rectified image maximum information of the original image. In other words, a pixel in the original image should be mapped to at least one pixel in the rectified image. Once the size is determined, the rectifying matrix (a homography) can be easily computed [1], and the color in the rectified image is computed through bilinear or bicubic interpolation from the original image.

White Balancing and Image Enhancement. The goal of color enhancement is to transform the input whiteboard image into an image with the same pen strokes on uniform background (usually

white). Many image enhancement techniques can be found at [2]. Here, we take advantage of knowledge that we are dealing with whiteboard images. Locally, whiteboard pixels are brighter than strokes. Our system computes the blank whiteboard image by inferring the value of pixels covered by the strokes from their neighbors. Rather than computing the blank whiteboard color at the input image resolution, our computation is done at a coarser level to lower the computational cost. This approach is reasonable because the blank whiteboard colors normally vary smoothly. Once the image of the blank whiteboard is computed, the input image is color enhanced in two steps:

1. Make the background uniformly white. For each cell, the computed whiteboard color is used to scale the color of each pixel in the cell.
2. Reduce image noise and increase color saturation of the pen strokes. We remap the value of each color channel of each pixel according to an S-shaped curve.

More details are to be found in our technical report [8].

3.2. Example

We have tested the proposed technique with more than 50 images taken by different people with different cameras in different rooms. All the tuning parameters have been fixed once for all. The success rate for automatic whiteboard detection is more than 90%. The four failures are due to poor boundary contrast, or to too noisy edge detection. Figure 2 shows a complete example of how our system works. The resolution is 2272×1704 pixels. As can be seen in the edge image (Fig. 2b), the actual lower border of the whiteboard does not have strong edge information, so the line corresponding to the pen holder is detected. The whiteboard corners estimated by intersecting the detected lines are shown in small red dots in Fig. 2a. The cropped and rectified image is shown in Fig. 2d. The estimated whiteboard color as if there were nothing written on it is shown in Fig. 2e, and the final enhanced image is shown in Fig. 2f.

4. WHITEBOARD SCANNING SUBSYSTEM

The major steps of the Whiteboard Scanning system is illustrated in Figure 3, and will be explained below. The mathematic foundation is that two images of a *planar* object, regardless the angle and position of the camera, are related by a plane perspectivity, represented by a 3×3 matrix called *homography* **H** [3, 1]. The stitching process is to determine the homography matrix between successive images, and we have developed an automatic and robust technique based on points of interest. This has several advantages over classical stitching techniques based on minimizing color differences: (1) less sensitive to color changes between images due to e.g. different focus; (2) less likely converge to local minima because those points of interest contain the most useful information and other textureless whiteboard pixels, which are distracting in color-based optimization, are discarded; (3) robust to large motion because a global search based on random sampling is used.

During whiteboard scanning, we start taking a snapshot from the upper left corner, a second by pointing to the right but having overlap with previous snapshot, and so on until reaching the upper right corner; move the camera lower and take a snapshot, then take another one by pointing to the left, and so on until reaching the left edge; the process continues in the "S" way until the lower border is captured. Successive snapshots must have overlap to allow later
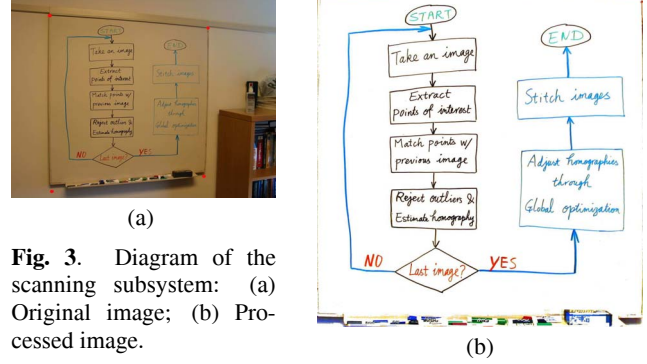

(a)

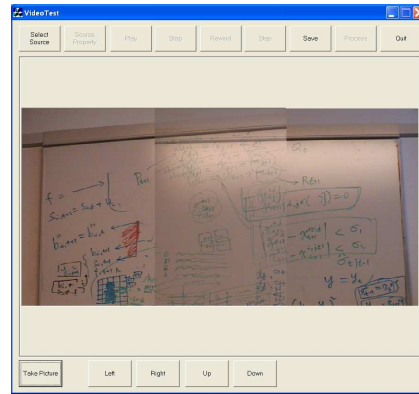**Fig. 3**. Diagram of the scanning subsystem: (a) Original image; (b) Processed image.

(b)

Fig. 4. User interface for whiteboard scanning. Note that half of the previous acquired image and half of the live video are shown in transparent to guide the user to take the next snapshot.

stitching, and this is assisted by providing visual feedback during acquisition, as shown in Figure 4. We suggest having half image to overlap between successive images. In the viewing region, we show both the previously acquired image and the current video view. In order to facilitate the image acquisition, half of the previously acquired image is shown in opaque, while the other half, which is in the overlapping region, is shown in semi-transparent. The current live video is also shown in half opaque and half semi-transparent. This guides the user to take successive images with overlap. Note that the alignment does not need to be precise. Our program will automatically align them. There are also a few buttons to indicate the direction in which the user wants to move the camera (down, up, left, right). The overlapping region changes depending on the direction. We have designed the default behavior such that only the "down" button is necessary to realize image acquisition in the "S" way.

Referring to Figure 1. For each image acquired, we use the Plessey corner detector, a well-known technique, to extract points of interest. These points correspond to high curvature points in the intensity surface if we view an image as a 3D surface with the third dimension being the intensity. An example is shown in Figure 5a, where the extracted points are displayed in red +.

Next, we try to match the extracted points with those from the previous image. For each point in the previous image, we choose an 15 × 15 window centered on it, and compare the window with windows of the same size, centered on the points in the current image. A zero-mean normalized cross correlation between two windows is computed. It ranges from -1, for two windows which are not similar at all, to 1, for two windows which are identical. If the largest correlation score exceeds a prefixed threshold (0.707 in our case), then that point in the current image is considered to be the match candidate of the point in the previous image. The match
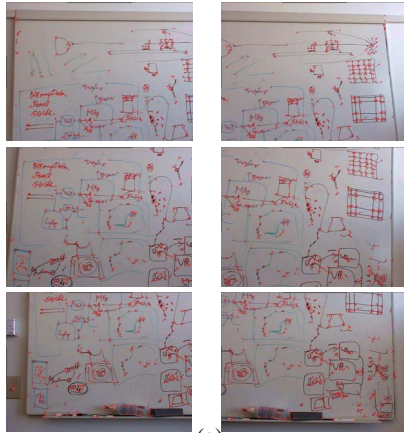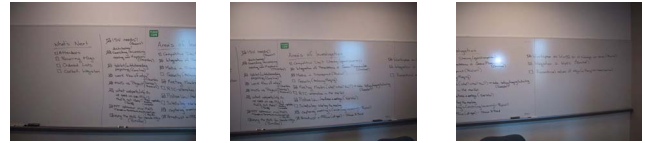
(a)
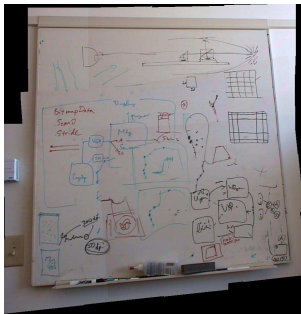
(b)
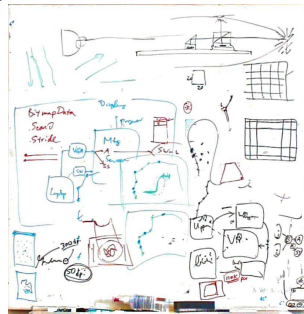
(c)

Fig. 5. An example of whiteboard scanning. (a) Original images overlayed with detected points of interest; (b) Stitched image; (c) Processed image using the technique described in the last section.

(a)

(b)

(c)

Fig. 6. A second example of whiteboard scanning. (a) Three original images; (b) Stitched image; (c) Final processed image.

candidate is retained as a match if and only if its match candidate in the previous image happens to be the point being considered. This symmetric test reduces many potential matching errors.
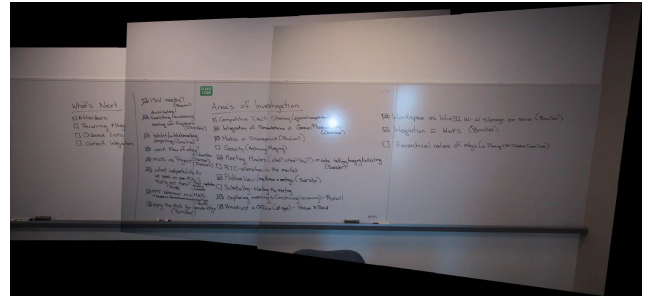
The set of matches established by correlation usually contains false matches because correlation is only a heuristic and only uses local information. Inaccurate location of extracted points because of intensity variation or lack of strong texture features is another source of error. The geometric constraint between two images is the homography constraint. If two points are correctly matched, they must satisfy this constraint, which is unknown in our case. If we estimate the homography between the two images based on a least-squares criterion, the result could be completely wrong even if there is only one false match. This is because least-squares is not robust to outliers. We developed a technique based on a robust estimation technique known as the *least median squares* (see e.g. [7]) to detect both false matches and poorly located corners, and simultaneously estimate the homography matrix $\mathbf{H}$.

This incremental matching procedure stops when all images have been processed. Because of incremental nature, cumulative errors are unavoidable. For higher accuracy, we need to adjust $\mathbf{H}$'s through global optimization by considering all the images simultaneously. Again, due to space limitation, the reader is referred to our technical report for details [8].
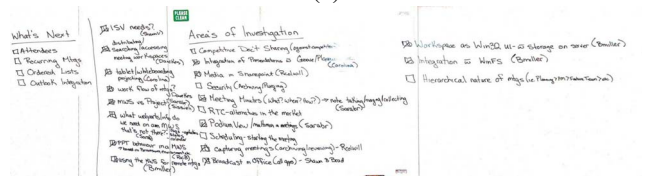
Once the geometric relationship between images (in terms of homography matrices $\mathbf{H}$'s) are determined, we are able to stitch all images as a single high-res image. There are several options, and currently we have implemented a very simple one. We use the first image as the reference frame of the final high-res image, and map successively original images to the reference frame. If a pixel in the reference frame appears several times in the original images, then the one in the newest image is retained.

Two examples are shown in Fig. 5 and Fig. 6.

5. CONCLUSIONS

We have presented a digital notetaking system by scanning the content on a whiteboard into the computer with a camera. Images are enhanced for better visual quality. The system has been tested extensively, and very good results have been obtained. Because digital cameras are becoming ubiquitous, our technology may contribute to a significant increase in productivity.

6. REFERENCES

[1] O. Faugeras and Q.-T. Luong, *The Geometry of Multiple Images*, MIT Press, 2001.

[2] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, 2nd edition, Prentice Hall, 2002.

[3] R. Hartley and A. Zisserman, *Multiple View Geometry*, Cambridge University Press, 1998.

[4] L. He, Z. Liu, and Z. Zhang, "Why take notes? use the whiteboard system," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, Hong Kong, Apr. 2003, vol. V, pp. 776–779.

[5] R. Jain, R. Kasturi, and B.G. Schunck, *Machine Vision*, McGraw-Hill, Inc., 1995.

[6] E. Saund, *Image Mosaicing and a Diagrammatic User Interface for an Office Whiteboard Scanner*. Technical Report, Xerox Palo Alto Research Center, 1999.

[7] Z. Zhang, "Parameter estimation techniques: a tutorial with application to conic fitting", *Image and Vision Computing*, 15(1):59–76, 1997.

[8] Z. Zhang and L. He, *Whiteboard Scanning and Image Enhancement*, Microsoft Research Technical Report, 2003. (In Press)