DIGITAL TECHNOLOGY FOR ORDINARY WHITEBOARDS

Zhengyou Zhang and Li-wei He

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

ABSTRACT

We will demonstrate 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.

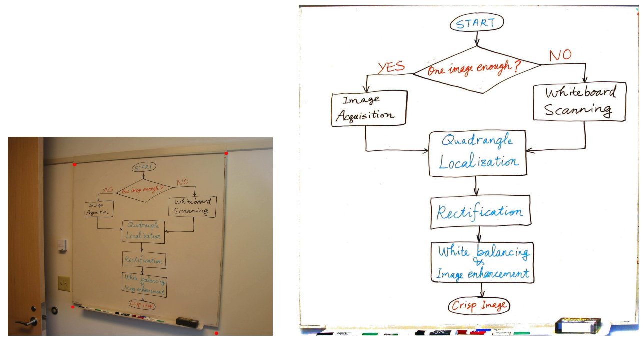
1. INTRODUCTION

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.

The reader can find full details in our technical report [1].

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 digital 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.



(a) (b)

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

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

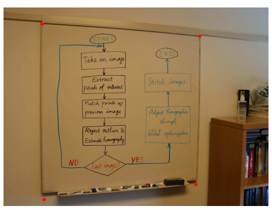
The first step is then to localize the borders of the whiteboard in the image. This is done by detecting four strong edges. The whiteboard in an image usually appears to be a general quadrangle, rather than a rectangle, because of camera’s perspective projection. If a whiteboard does not have strong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

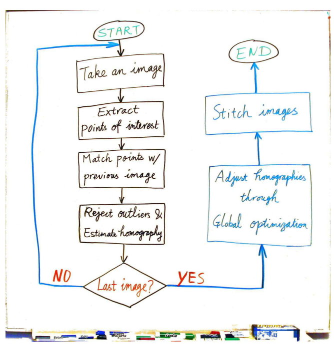
The last step is white balancing of the background color. This involves two procedures. The first is the estimation of the background color (the whiteboard color under the same lighting without anything written on it). This is not a trivial task because of complex lighting environment, whiteboard reflection 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. WHITEBOARD SCANNING SUBSYSTEM

The major steps of the Whiteboard Scanning system is illustrated in Figure 2, 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* $\mathbf{H}$. 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.



(a)



(b)

Fig. 2. Diagram of the scanning subsystem: (a) Original image; (b) Processed image.

Referring to Figure 2. 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.

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 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* to detect both false matches and poorly located corners, and simultaneously estimate the homography matrix $\mathbf{H}$.

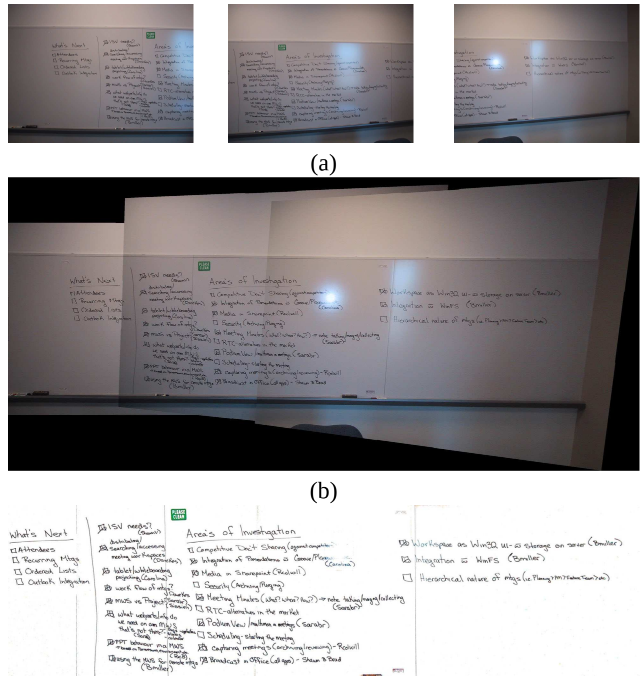


Fig. 3. An example of whiteboard scanning. (a) Three original images; (b) Stitched image; (c) Final processed image.

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 [1].

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.

An example is shown in Fig. 3.

4. REFERENCES

[1] Z. Zhang and L. He, *Whiteboard Scanning and Image Enhancement*, Microsoft Research Technical Report TR-03-39, 2003.