Intelligent Document Capturing and Blending System Based on Robust Feature Matching with an Active Camera

Wan-Yu Chen, Jia-Lin Chen, Yu-Chi Su, and Liang-Gee Chen, *Fellow, IEEE*
DSP/IC Design Lab., Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan

***Abstract*-- We propose an intelligent document capturing and blending system based on robust feature matching for efficient document management. The proposed system not only supports handwritten text and figure extraction, but also provides image blending mechanism to automatically merge the extracted handwritten texts and figures into electronic documents for the user. The proposed system addresses camera shake and luminance variation problems caused by active cameras. Besides, we adopt robust feature matching techniques to improve the system accuracy. Experimental results show that our system supports 95.65% detection rate and achieves 88.3% compression ratio reduction compared with the previous work. Besides, we also compare system performance considering Scale Invariant Feature Transform (SIFT) [1] and Speeded-Up Robust Features (SURF) [2]. We derive 71.2% complexity reduction and 4.3% detection rate degradation with SURF feature matching.**

I. INTRODUCTION

Recently, because of the facility of information indexing, retrieval, storage and exchange, documents in electronic form are more and more popular. However, a lot of paper documents still exist in our daily life. For example, when attending a meeting or course, we usually print the presentation documents and write some notes or comments on the paper documents. Such behavior helps us to understand and recall the presentation easily. In addition, we usually print technical papers and write some innovation idea on the papers when we read them. This action helps us to record our instant idea immediately and conveniently. Hence such handwritten record is valuable and needs efficient management. However, paper documents consume large storage space. And it is time-consuming to search a note from a lot of paper documents. The inconvenient management of paper documents makes our valuable idea hard to be maintained.

Along with the development of hand-held camera and scanner, people choose to convert paper documents into electronic form. Traditional flatbed scanner has better image quality but users usually feel cumbersome to use it. Hand-held camera is more popular than flatbed scanner. People usually carry a cellphone with camera every day and can capture a photo anytime. While the image sensor converts the paper documents into electronic image files, the captured image is constituted with image pixels without registration. And the image size is proportional to image quality. Several previous works have been presented to register the captured image for active camera applications [3]-[5]. Park et al. developed a vertical line detection method for image registration [3]. Kim et al. adopted rectangle line feature detection method using Hough transform [4]. Yuan et al. proposed a robust feature matching approach with pre-defined layout for better registration quality [5]. However, such capturing approach only assumes the document with the pre-defined layout and cannot adapt to dedicated document layout. Besides, the previous works only reduces the storage of paper document, users still feel inconvenient to search item from a lot of image files.

In this paper, we propose a novel document capturing and blending system for efficient document management. With robust feature matching over the original electronic documents, we successfully register the captured images with dedicated layout. Furthermore, the handwritten texts and figures are extracted and blended with original electronic documents by the proposed texture extraction and blending mechanism. The system aims to provide the user key-word search function of electronic documents and keep the handwritten texts and figures simultaneously. Users can share their handwritten record easily with cloud and local device cooperation. Besides, we address registration instability problems with an active camera and improve the registration quality by matching the dedicated layout. Furthermore, the compression ratio can be largely reduced due to handwritten texture extraction. Finally, we speed up the whole system with SURF feature matching with acceptable accuracy degradation.

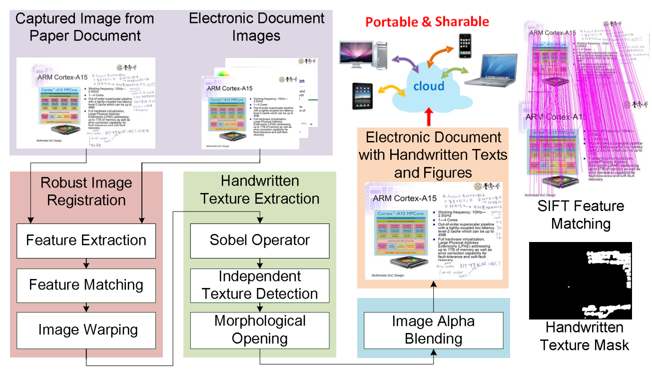


Fig. 1 System block diagram

II. PROPOSED METHOD

The overall block diagram of the system is shown in Fig. 1. The captured image is processed through feature extraction and matching to derive the global camera motion for robust image registration. The following is an image warping process to register the captured document based on estimated camera homography matrix. Second, we propose a handwritten texture extraction mechanism considering both original electronic document image and transformed captured image. Finally, we apply image alpha blending with the extracted handwritten texture mask on original electronic image to preserve handwritten texts and figures for the user.

A. Robust Image Registration

To register the captured image in a dynamic environment, a robust feature extraction and matching approach is employed as the first step of the framework. For each image, global camera homography matrix needs to be estimated before handwritten texture extraction. This is because the global camera motion mounted on a mobile agent differs frame by frame. By taking the advantage of robust feature extraction and matching techniques, we match corresponding feature points of the captured image with the original electronic document regardless of environment variation. To complete robust image registration, we separate the process into three steps. Firstly, SIFT features [1], a local descriptor with good scale, rotation, and luminance invariance, are extracted from the input image. Next, for each feature in the input image, feature matching is performed to find the nearest neighbors among reference images in the database. To speed up the processing time of the matching stage, kd-tree [6] is adopted as the index of the image database. After this step, hundreds of matching pairs for each frame in the input video are obtained. In addition, to remove false matching among matching pairs, RANSAC [6] is employed to filter outlier pairs. RANSAC algorithm iteratively selects samples at random among matching pairs and estimates their

978-1-4673-1363-6/13/$31.00 ©2013 Crown

homography matrix as the fitting model. Finally, we apply image warping using the estimated homography matrix H on the captured image as equation (1). $(\tilde{x},\tilde{y})$ is image coordinate after image warping.

$$H=\begin{bmatrix} h11 & h12 & h13 \\ h21 & h22 & h23 \\ h31 & h32 & h33 \end{bmatrix} \qquad \begin{bmatrix} w*\tilde{x} \\ w*\tilde{y} \\ w \end{bmatrix} = \mathrm{H}*\begin{bmatrix} \mathrm{x} \\ \mathrm{y} \\ \mathrm{z} \end{bmatrix} \tag{1}$$

B. Handwritten Texture Extraction

A Sobel filter-based texture detection technique combined with morphological opening operation is used to provide robust handwritten texture extraction. We design an independent Sobel filter-based mechanism to extract the texture area regardless of luminance and color variance between captured image and electronic document image. The idea of Sobel filter-based mechanism comes from the fact that handwritten texture detection reveals its strong edge response on captured image and weak edge response on original electronic document image. To reduce the image blur effect caused by homography transform, 10x10 window calculation is adopted in the proposed system. In this situation, handwritten texture mask is extracted more accurately.

Equation (2) shows Sobel filter operation. Equation (3) derives independent texture extraction based on the transformed paper document image *f(x,y)* and electronic document image *f'(x,y)*. On the other hand, even if most handwritten texture can be successfully extracted by the independent texture detection step, it usually fails to distinguish noise of captured image from handwritten texture, especially under severe camera noise. In this case, morphological opening operation is adopted to reduce handwritten texture misdetection caused by camera noise.

$$\begin{aligned} G_x &= f(x-1,y-1)+2f(x-1,y)+f(x-1,y+1))-(f(x+1,y-1)+2f(x+1,y)+f(x+1,y+1)) \\ G_y &= f(x-1,y-1)+2f(x,y-1)+f(x+1,y-1))-(f(x-1,y+1)+2f(x,y+1)+f(x+1,y+1)) \\ Sobel\,(f(x,y)) &= |G_x|+|G_y| \end{aligned} \tag{2}$$

$$\begin{aligned} & If\,((\sum_{(i,j)\in N} Sobel\,(f'(x+i,y+j)) < th) == 0)\ \&\&\ (Sobel\,(f(x+i,y+j)) > th) \\ & \quad texture\,(x,y) = true \\ & else \\ & \quad texture\,(x,y) = false \end{aligned} \tag{3}$$

C. Image Alpha Blending

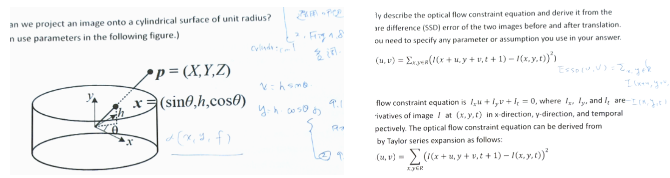
The image alpha blending is applied on the original electronic document with the handwritten texture mask extracted from the previous steps. Thus handwritten texts and figures are blended with the electronic document according to the texture mask.

When all the procedures illustrated above are done, the next captured image is processed by the system in a similar way. Most importantly, handwritten texture extraction and blending in this work are employed to provide both the convenient management and instant recording facility of electronic and paper documents for users.

III. EXPERIMENTAL RESULTS AND ANALYSIS

Two experiments are conducted to evaluate performance of the proposed system. All the images tested in experiments are captured from a CMOS front-mounted Canon S100 camera with 640×480 resolution.

Table I shows the system improvement on detection rate involving SIFT or SURF image registration mechanism, respectively. With the proposed image registration mechanism, the detection rate is improved to 95.65% and 91.3% considering SIFT and SURF techniques. Fig. 2 depicts the success and failure case of SURF based image registration. From Fig.2 (a), we can see that an image with figures could be easily detected by SURF. But an image containing a lot of texts and no distinguished layout, as shown in Fig. 2 (b), is difficult for SURF matching. In this case, the SIFT technique can distinguish the feature difference of individual texts and achieves higher accuracy.



(a) (b)

Fig. 2 Results of success and failure images. (a) Success image with SURF approach. (b) Failure image with SURF approach.

TABLE I ACCURACY COMPARISON

Detection method	Accuracy (%)
SIFT feature matching	95.65
SURF feature matching	91.30

The second experiment evaluates the compression ratio of the proposed system. In the proposed system, user only needs to store the handwritten texture area and reduce the disk space. This experimental result demonstrates that our system provides 88.3% compression ratio reduction compared with the previous method [4].

TABLE II TIME COMPLEXITY ANALYSIS

Module	SIFT(%)	SURF(%)
Feature Extraction	88.64	74.39
Feature Matching	5.08	0.95
Image Warping	1.09	3.79
Others	6.09	21.07

Table II summaries the workload ratio of each module by software implementation for the proposed system. The operating environment is under Win7 operating system with Intel Core I7 3.4G CPU and 4GB DDR RAM. This result shows feature extraction of robust image registration dominate 88.64% of the whole process. The system process takes 5.29 second in average to deal with a frame with SIFT techniques. To speed up the system performance, we adopt SURF feature extraction. Thus, our system is accelerated to 1.53 second in average per frame and this processing speed is acceptable by user experience.

IV. CONCLUSION

We propose an efficient document capturing and blending system based on robust feature matching for active camera applications. With the proposed system, we can digitalize handwritten texts and figures on the paper documents and record it on electronic documents automatically. Thus, we can exploit both the instant recording advantages of paper documents and convenient management of electronic documents.

Our system achieves 95.65% detection rate with SIFT feature matching technique and 88.3% compression ratio reduction compared with the previous work [4]. In addition, we also derive 71.2% complexity reduction and 4.3% detection rate degradation with SURF features.

REFERENCES

[1] D. G. Lowe,"Distinctive image features from scale-invariant keypoints", published by International Journal of Computer Vision, 2004.

[2] H. Bay, et al., "SURF: Speeded Up Robust Features", ECCV, Vol. 110, pp. 407-417, 2006.

[3] A. Park, et al., "Intelligent document scanning with active camera", ICDAR, pp. 991 – 995, vol. 2, 29 Aug.-1 Sept. 2005.

[4] W. H. Kim, et al., "Document Capturing Method with a Camera Using Robust Feature Points Detection", DICTA, pp.678 – 682, Dec. 2011.

[5] V. G. Edupuganti, et al., "Registration of camera captured documents under non-rigid deformation", CVPR, pp. 385-392, 2011.

[6] N. Y. Khan, et al., "Performance Evaluation against Various Image Deformations on Benchmark Dataset", DICTA, pp.503 – 506, Dec. 2011.