AUTOMATIC BUSINESS CARD SCANNING WITH A CAMERA

Gang Hua[†], Zicheng Liu[‡], Zhengyou Zhang[‡] and Ying Wu[†]

[†]2145 Sheridan Road, EECS Dept.
Northwestern University
Evanston, IL 60208, U.S.A.
{ganghua, yingwu}@ece.northwestern.edu

[‡]One Microsoft Way
Microsoft Research
Redmond, WA 98052, U.S.A.
{zliu, zhang}@microsoft.com

ABSTRACT

In this paper, we present a system to automatically extract, rectify and enhance business card images. First the business card image patch is automatically segmented by minimizing a novel local-global variational energy. Second a quadrangle is fitted to the segmented image patch. With the four corner points of the quadrangle, we then estimate the physical aspect ratio of the business card and obtain a homography to rectify the quadrangle back to rectangular shape. We finally enhance the contrast of the rectified business card image using a S-shaped curve. Extensive experiments demonstrated the efficacy and robustness of our system.

1. INTRODUCTION

People use business cards a lot. In professional meetings, the exchange of business cards is a common way for professionals to build connections. As the number of business cards grows, organizing them becomes a burden. Realizing that the most effective way of organizing business card information is to digitize them, many researchers have developed systems to use flatbed scanners to scan business cards and use OCR technology to convert to digital. One drawback of this approach is that people usually do not have flatbed scanners at hand. As more and more mobile phones are equipped with cameras and the camera resolution keeps improving, mobile phones are becoming convenient devices for business card scanning and digitization. Compared with flatbed scanners, mobile devices present new technical challenges for image processing. First, the lighting condition is unpredictable. Second, the orientation of the business card can be arbitrary. Third, the background may be cluttered. Therefore, before feeding an image to an OCR system, we need to extract the business card, rectify and enhance it. This is the focus of this paper.

Previous work for general scene text analysis [8, 2, 3, 1, 7, 5] detect each scene text character or word directly from the images. The extracted texts are usually subject to perspective distortion which needs to be rectified before sending the texts to the back-end systems for further processing. One common

---

approach is to use the principal orientation of the group of texts to rectify the texts [1, 8]. This approach is attractive. One drawback, though, is that it is difficult to handle the situation where there are texts on the background. It would be difficult to determine which texts belong to the business card and which belong to the background. In addition, the background texts may affect the estimation of the principal text orientation resulting in inaccurate rectification of the texts.

This challenge may be greatly relieved if we can correctly identify the boundary of the business card. A business card is usually rectangular and it will be mapped to a quadrangle in the images. If we rectify the quadrangle back to a rectangle with the correct aspect ratio, the texts on the business card will be rectified simultaneously. Therefore, our business card scanning system includes four processing steps:

1. The automatic segmentation of the business card image region is performed by minimizing a novel local-global variational energy [4]. This will obtain the boundary chain code of the business card image region and the binary masks of both the business card and background regions. We summarize the algorithm in Sec. 2.

2. Then, we fit a quadrangle on the business card image region. This step will identify the four corner points of the quadrangle shaped business card region. We describe the details in Sec. 3.

3. With the four corner points, we estimate the physical aspect ratio of the business cards using the approach similar to that in [9], and thus rectify it to be rectangular by a homography. The contents are described in Sec. 4.

4. The contrast of the rectified business card image is then enhanced by transform the pixel values through a S-shaped curve, which is presented in Sec. 5.

In this paper, we make the assumption that the business card is at the focus of attention, i.e., it is roughly at the image center. This assumption is reasonable because the user will try to put the business card in the image center when he takes the picture with his mobile phone. Notice that this does not directly tell us where the business card boundary is, but it facil-

itates the automatic initialization of the business card region segmentation algorithm and thus automates all the aforementioned processing steps. Our system runs in real time and we present extensive experimental results in Sec. 6. Finally we conclude in Sec. 7.

2. BUSINESS CARD SEGMENTATION

We denote $\mathcal{A}_\mathcal{F}$ and $\mathcal{A}_\mathcal{B}$ as the business card and background image regions, thus the business card boundary is $\Gamma(c) = \mathcal{A}_\mathcal{F} \cap \mathcal{A}_\mathcal{B}$ and the image data $\mathcal{I} = \mathcal{A}_\mathcal{F} \cup \mathcal{A}_\mathcal{B}$. We model the coherence of both the business card region and background region probabilistically using two GMMs $P_\mathcal{F}(\mathbf{u})$ and $P_\mathcal{B}(\mathbf{u})$ respectively, where $\mathbf{u}$ is the 5 dimensional feature vector consists the LUV pixel value and the (x, y) pixel location. The image data can thus be modeled by a generative mixture model $P_\mathcal{I}(\mathbf{u}) = \omega_\mathcal{F} P_\mathcal{F}(\mathbf{u}) + \omega_\mathcal{B} P_\mathcal{B}(\mathbf{u})$, where $\omega_\mathcal{F}$ and $\omega_\mathcal{B}$ are the mixture weights.

The segmentation is then performed by minimizing the following variational energy [4]

$$\mathbf{E}_p(\Gamma(c), P_\mathcal{I}) = \alpha \mathbf{E}_e - \beta \mathbf{E}_h - \gamma \mathbf{E}_l \qquad (1)$$

where

$$\mathbf{E}_e = \int_0^1 \frac{1}{1 + |\mathbf{g}_x(\Gamma(c))| + |\mathbf{g}_y(\Gamma(c))|} |\dot{\Gamma}(c)| dc \qquad (2)$$

$$\mathbf{E}_h = \int_{\mathcal{A}_\mathcal{F}} \log \{\omega_\mathcal{F} P_\mathcal{F}(\mathbf{u})\} + \int_{\mathcal{A}_\mathcal{B}} \log \{\omega_\mathcal{B} P_\mathcal{B}(\mathbf{u})\} \qquad (3)$$

$$\mathbf{E}_l = \int_{\mathcal{A}_\mathcal{F} \cup \mathcal{A}_\mathcal{B}} \log\{\omega_\mathcal{F} P_\mathcal{F}(\mathbf{u}) + \omega_\mathcal{B} P_\mathcal{B}(\mathbf{u})\}, \qquad (4)$$

with $\mathbf{g}_x$ and $\mathbf{g}_y$ being the image gradient and $\alpha + \beta + \gamma = 1$. We call $\mathbf{E}_e$, $\mathbf{E}_h$ and $\mathbf{E}_l$ the boundary potential, local region potential and global likelihood potential, respectively.

The $\mathbf{E}_p(\Gamma(c), P_\mathcal{I})$ is minimized by a sequential iterative optimization scheme: fixing $P_\mathcal{I}$, optimizing the $\Gamma(c)$ by level set; and fixing $\Gamma(c)$, re-estimate $P_\mathcal{I}$ by fixed-point iteration. These two steps guarantee to decrease $\mathbf{E}_p$ and it can be automatically initialized based on the focus-of-attention assumption. Due to the page limit, we refer to [4] for more detailed discussion of the local-global energy formulation, as well as the details of the optimization scheme. Indeed, compared with previous work [10, 6], the uniqueness of the local-global energy formulation Eq. 1 resides on the global image likelihood $\mathbf{E}_l$, which makes the model estimation step more accurate when $\mathcal{A}_\mathcal{F}$ and $\mathcal{A}_\mathcal{B}$ inaccurately contains portion of the pixels of the true background and foreground, respectively.

3. BUSINESS CARD QUADRANGLE FITTING

The physical shape of a business card is usually rectangle. However, it will be projected as a quadrangle shape in the image, which skews the texts. If we transform the quadrangle back to a rectangle with the correct aspect ratio, we rectify the skewness of the texts simultaneously. Since the segmented card region is not a perfect quadrangle due to occlusion, background clutter, etc.. We need to fit a quadrangle on it and this is formulated as an optimization problem.

The output from the segmentation algorithm is the business card boundary chain code, which is a dense polygon representation. We denote $\mathbf{V} = \{\mathbf{v}_0, \ldots, \mathbf{v}_{n-1}\}$ as the set of n dense vertices, and denote $\mathcal{L} = \{l_0, \ldots, l_{r-1}\}$ be the $r = C_n^2$ line sets formed by all lines $\overline{\mathbf{v}_i \mathbf{v}_j}$, where $i \neq j$ and $0 \leq i, j \leq n - 1$. The set of all s quadrangles formed by four different lines in $\mathcal{L}$ is denoted by $\mathcal{Q} = \{Q_0, \ldots, Q_{s-1}\}$ where $s = C_r^4$. We also denote $\{\theta^0, \theta^1, \theta^2, \theta^3\}$ as the four corner angles of quadrangle $Q \in \mathcal{Q}$. Let N_F, N_Q and $N_{F \cap Q}$ be the three numbers of pixels in the segmented card region, inside Q and in the intersection of the former two, respectively. Also let n_c be the number of vertices in $\mathbf{V}$ which are in the 3×3 neighborhood of the boundary line pixels of Q. The fitness of the Q to the segmented region is defined as

$$\mathcal{S}(Q) = \frac{n_c}{n} \sqrt[4]{\prod_{w=0}^3 (1 - |\cos \theta^w|)} \sqrt{\frac{N_{F \cap Q}}{N_Q} + \frac{N_{F \cap Q}}{N_F}}. \qquad (5)$$

Therefore, the quadrangle fitting is just to solve for an optimization problem which seeks for the best Q^* such that

$$Q^* = \arg \max_{Q \in \mathcal{Q}} \mathcal{S}(Q). \qquad (6)$$

In principle, it favors quadrangles whose boundary and enclosed region coincide the most with the segmented business card image region. It also favors quadrangles whose corner angles are near $\frac{\pi}{2}$. This is based on the assumption that the users try to face the front of the business card to the camera.

However, when $n = 300$ (typical for $\mathbf{V}$), the cardinality of $\mathcal{Q}$ is $s = C_r^4 = 4.024e + 016$. It is infeasible to exhaustively search in $\mathcal{Q}$ for Q^*. We develop the following approach to pruning the solution space, and thus obtain the Q^* more efficiently.

Curve Simplification: Indeed $\mathbf{V}$ is a highly redundant shape representation and can be simplified with high accuracy by:
● `Multiscale corner point detection`: Denote $(i)_m = i \, mod \, m$, then for $i = 0, \ldots, n-1$, check whether

$$\left| \frac{(\mathbf{v}_{(i-j)_n} - \mathbf{v}_i) \cdot (\mathbf{v}_{(i+j)_n} - \mathbf{v}_i)}{\|\mathbf{v}_{(i-j)_n} - \mathbf{v}_i\| \|\mathbf{v}_{(i+j)_n} - \mathbf{v}_i\|} \right| < 0.98 = \cos(10^o) \quad (7)$$

are satisfied for all $j = 1, \ldots, q$ (e.g., $q = 20$). If yes, we keep $\mathbf{v}_i$, otherwise we remove it. This step in principle removes vertices with too small transitions over multiple scales. We denote the reduced m vertices set as $\tilde{\mathbf{V}} = \{\tilde{\mathbf{v}}_0, \tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_{m-1}\}$.
● `Iterative minimum distance pruning`: For $i = 0, \ldots, m-1$, evaluate $d_i = d(\tilde{\mathbf{v}}_i, \overline{\tilde{\mathbf{v}}_{(i-1)_m} \tilde{\mathbf{v}}_{(i+1)_m}})$ the Euclidean distance from $\tilde{\mathbf{v}}_i$ to the straight line formed by its neighbor vertices $\tilde{\mathbf{v}}_{(i-1)_m}$ and $\tilde{\mathbf{v}}_{(i+1)_m}$. Suppose $\tilde{\mathbf{v}}_k$ is such that $d_k = \min_i \{d_i\}$, if $d_k < \epsilon_d$, where ϵ_d is a pre-specified tolerance (e.g., $\epsilon_d = 1$), we remove $\tilde{\mathbf{v}}_k$ from $\tilde{\mathbf{V}}$. Repeat the same

operations until no more vertices could be removed. This returns the final reduced l vertices set $\hat{V} = \{\hat{\mathbf{v}}_0, \ldots, \hat{\mathbf{v}}_{l-1}\}$.

Quadrangle pruning: With the pruned vertices set $\hat{V}$, we further prune the number of lines and quadrangles:

• `Ordered Lines`: For $0 \leq i \leq l - 1$, build the ordered line set $\mathbf{L}_i = \{\overline{\hat{\mathbf{v}}_i \hat{\mathbf{v}}_{(i+1)_l}}, \ldots, \overline{\hat{\mathbf{v}}_i \hat{\mathbf{v}}_{(i+n_d)_l}}\} = \{l_{i1}, \ldots, l_{in_d}\}$ where n_d specifies how far (e.g., $n_d = 4$) to look forward to form the lines. We obtain the ordered set $\mathbf{L} = \{\mathbf{L}_1, \ldots, \mathbf{L}_l\} = \{l_0, \ldots, l_{p-1}\}$. Notice that placing the order constraint reduces the number of lines from C_l^2 to $l \cdot n_d$.

• `Quadrangle searching`: Let $\mathbf{Q} = \{Q_{ijkf} : 0 \leq i < j < k < f \leq p - 1\}$ be all the possible quadrangles Q_{ijkf} spanned by $\{l_i, l_j, l_k, l_f\}$. Notice that not all quaternion $\{l_i, l_j, l_k, l_f\}$ can span a quadrangle. We then exhaustively search for the best quadrangle Q^* on $\mathbf{Q}$, i.e., $Q^* = \arg\max_{Q \in \mathbf{Q}} \mathcal{S}(Q)$. Moreover, a post-processing is performed, by first collecting the Sobel edge points in the neighborhood of the boundary line pixel of Q^* and then performing a weighted least square fitting to further refine the position of each side line of Q^*.

• `More heuristics`: The most intensive computation is to evaluate $\sqrt{\frac{N_{F \cap Q}}{N_Q} + \frac{N_{F \cap Q}}{N_F}}$ in $\mathcal{S}(Q)$, since we must count the number of pixels in the intersection of two image regions. The following criterions have been proven to be very effective to reduce the computation without losing much accuracy:

    ○ If the segment length of $\overline{\hat{\mathbf{v}}_i \hat{\mathbf{v}}_{(i+j)_l}}$, $1 \leq j \leq n_d$ is less than $\frac{1}{16}$ of the minimum of the image width and length, then we do not put it in $\mathbf{L}_i$.

    ○ If any corner point of Q_{ijkf} falls out of the size of the image, we simply discard it.

    ○ If $\frac{n_c}{n} < 0.5$ for Q_{ijkf}, we simply discard it.

    ○ If $|\cos \theta^w| > 0.2$ for any $w = 0, \ldots, 3$, the quadrangle Q is discarded without further evaluation.

4. BUSINESS CARD RECTIFICATION

Once we have obtained the best quadrangle Q^* for the business card in the image, we can easily identify the four corner points. Then by utilizing the approach similar to that in [9], we estimate the physical aspect ratio $R_\alpha = \frac{W_l}{W_h}$ of the business card. Notice that the aspect ratio estimation algorithm presented in [9] made some weak assumptions about the parameters of the cameras, which indeed can be satisfied by most of the cameras nowadays. We refer to [9] for more details due to the page limit.

To rectify the quadrangle back to rectangular shape, we need to determine the size of the rectangle after rectification. Since we do not want to lose any image information, i.e., each image pixel inside the quadrangle must have a direct map in the rectified rectangular image, we must set the length L_r and width W_r of the rectified rectangle to be suitable to achieve
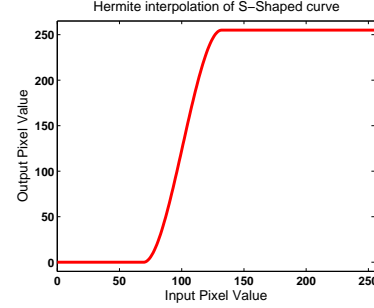

Fig. 1. An example of S-Shaped curve.

this. We firstly identify the longest side of the quadrangle whose length is denoted as L_q. We also denote the longer length of the two neighbor sides of the longest side as W_q. Then, if $R_\alpha W_q > L_q$, we set $W_r = W_q$ and $L_r = R_\alpha W_q$, otherwise we set $L_r = L_q$ and $W_r = \frac{L_r}{R_\alpha}$. We thus obtain the four corner points of the rectified rectangle to be $(0, 0)$, $(L_r, 0)$, (L_r, W_r) and $(0, W_r)$ and they are one-to-one corresponding to the four corner points of the quadrangle.

Since the business card is a planar object, a homography transform is sufficient to rectify it back to be a rectangular shape. It is obvious that the parameters of the homography transform can be conveniently estimated from the four pairs of corresponding points. We use it to re-warp the quadrangle shaped image patch to the rectified rectangle by reverse mapping with bi-linear color pixel interpolation.

5. BUSINESS CARD ENHANCEMENT

To make the contrast of the text characters and the background in the rectified business card image more sharp, we simply independently transform the R, G, B pixel values of the rectified image through a S-shape curve, which is obtained by Hermite polynomial interpolation on the average intensity $\bar{\mathcal{L}}_l$ of the lightest 10% pixels and the average intensity $\bar{\mathcal{L}}_d$ of the darkest 10% pixels. In principle, the curve should map the pixel value larger or equal to $\bar{\mathcal{L}}_l$ and pixel value less or equal to $\bar{\mathcal{L}}_d$ to near 255 and 0, respectively. We present in Fig. 1 an example S-shaped curve interpolated in our experiments.

6. EXPERIMENTS

We have tested our system extensively and quite good results have been obtained. For example, we have tested our system on over 300 live captured business card images and the overall successful rate is about 95%. The success or failure of one test is evaluated subjectively based on the manually segmented results. Fig. 2 presents some of the example results from our Business card scanning system. The first row to the fifth row show the results from each of the processing steps of our system, respectively. In the first and last column, it clearly shows that the background texts does not affect the rectification results at all. Also notice how the quadrangle fitting step identified the correct quadrangle even when some

(a). Card segmentation results, boundary curve is shown in red.

(b). Results of curve simplification, white points constitute the simplified vertex set.
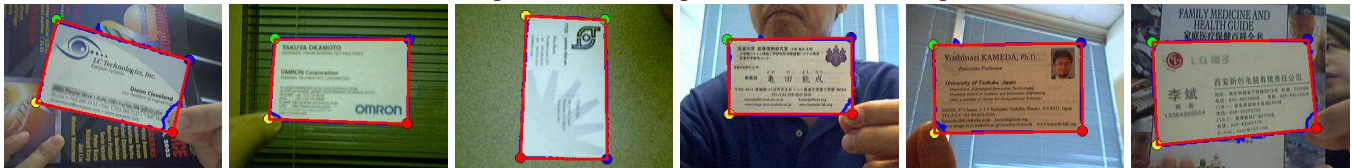
(c). Quadrangle fitting of business card, the four corner points are shown in yellow, green, blue and red.

(d). Results of rectified business card.

(e). Results of enhanced business card image.

Fig. 2. Business card scanning results: First row: Card Segmentation. Second row: Curve Simplification. Third row: Quadrangle Fitting. Fourth row: Shape Rectification. Fifth row: Card Enhancement.

corner points are occluded by human fingers. Our system is quite robust to both background texts and partial occlusions.

7. CONCLUSION AND FUTURE WORK

We have presented a fully automatic business card scanning system with cameras, which targets at the application of mobile assistant. We have tested the system extensively and obtained good results. In the future, we are planning on integrating our system with an OCR package and developing a complete business card digitization system on mobile devices.

8. REFERENCES

[1] Y. Cui and Q. Huang. Character extraction of license plates from video. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 502–507, San Juan, Puerto Rico, June 1997.

[2] J. Gao and J. Yang. An adaptive algorithm for text detection from natural scenes. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 84–89, Kauai Marriott, Hawaii, December 2001.

[3] I. Haritaoglu. Scene text extraction and translation for handheld devices. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 408–413, Kauai Marriott, Hawaii, December 2001.

[4] G. Hua, Z. Liu, Z. Zhang, and Y. Wu. Iterative local-global energy minimization for automatic extraction of object of interest. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2005. Submitted.

[5] A. K. Jain and B. Yu. Automatic text location in images and video frames. *Pattern Recognition*, 31(22):2055–2076, December 1998.

[6] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, pages 223–247, 2002.

[7] V. Wu, R. Manmatha, and E. M. Riseman. Textfinder: An automatic system to detect and recognize text in images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1224–1229, November 1999.

[8] J. Yang, J. Gao, Y. Zhang, X. Chen, and A. Waibel. An automatic sign recognition and translation system. In *Workshop on Perceptual User Interfaces*, November 2001.

[9] Z. Zhang and L. He. Notetaking with a camera: Whiteboard scanning and image enhancement. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 533–536, Montreal, Quebec, Canada, May 2004.

[10] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transaction on Pattern Recognition and Machine Intelligence*, 18(9):884–900, 9 1996.