Robust Vanishing Point Detection for MobileCam-Based Documents

Xu-Cheng Yin, Hong-Wei Hao
Department of Computer Science
School of Computer and Communication Engineering,
University of Science and Technology Beijing
Beijing, China
xuchengyin@ustb.edu.cn; hhw@ustb.edu.cn

Jun Sun, Satoshi Naoi
Fujitsu R&D Center Co. Ltd.
Beijing, China
sunjun@cn.fujitsu.com; naoi@cn.fujitsu.com

***Abstract*—Document images captured by a mobile phone camera often have perspective distortions. In this paper, fast and robust vanishing point detection methods for such perspective documents are presented. Most of previous methods are either slow or unstable. Based on robust detection of text baselines and character tilt orientations, our proposed technology is fast and robust with the following features: (1) quick detection of vanishing point candidates by clustering and voting on the Gaussian sphere space; and (2) precise and efficient detection of the final vanishing points using a hybrid approach, which combines the results from clustering and projection analysis. The rectified image acceptance rate for MobileCam-based documents, signboards and posters is more than 98% with an average speed of about 100ms.**

Keywords-Vanishing point detection, Perspective document rectification, clustering, MobileCam-based documents, the Gaussian sphere

I. INTRODUCTION

With widespread usage of the cheap digital camera built-in the mobile phone (MobileCam in abbreviation thereafter) in people's daily life, the demand for simple, instantaneous capture of document images emerges. Different from the traditional scanned image, lots of the MobileCam-based document images have perspective distortions. Consequently, rectifying MobileCam-based perspective document images becomes an important issue [6][10].

In computer vision, most perspective correction methods rely on vanishing point detection. And these methods involve extracting multiple lines and their intersections, or using texture and frequency knowledge [4][11][17][19]. In document analysis, there are also various works on correction of perspective documents captured by general digital cameras. We divided the methods of vanishing point detection into two sub groups: direct approaches and indirect approaches. The direct approaches directly analyze and calculate on image pixels, such as projection analysis from a perspective view for horizontal vanishing point detection [2]. These approaches have rather good precisions. But a full or partial search of the space is computationally expensive, even impossible. The indirect approaches convert the original space into a clue space, and search the vanishing point in that new small space. Most indirect approaches involve extracting multiple straight or illusory lines and voting vanishing points by model fitting [2][7][8][9][12][13]. These indirect approaches are time efficient. However, the model fitting is sensitive.

In MobileCam-based document analysis, there are two main challenges for rectifying perspective documents. First, the rectifying engine should be highly precise with fast speed. The above methods can't cover the two issues well at the same time. Second, a MobileCam-based image is usually a partial portion of a whole document with few document boundaries. To solve the above problems aiming at a practical MobileCam application, we focus on fast and robust vanishing point methods for rectifying general perspective documents. First, we propose a hybrid approach for robust real-time vanishing point detection by integrating the direct and indirect approaches efficiently. As for the second challenge, we utilize horizontal text baselines and character tilt orientations as illusory lines to vote and compute vanishing points.

The remainder of this paper is organized as follows. Section 2 introduces the basic principle of our vanishing point detection methods. In Section 3, we describe our vanishing point detection process with a hybrid approach. Section 4 is about the experiments and result analysis. Finally we conclude the paper in Section 5.

II. BASIC PRINCIPLE OF OUR METHODS

The vanishing point (horizontal or vertical) in a 2D space can be described as (vx, vy). Generally speaking, vanishing point detection is to find a proper point according to an optimization process in the image plane. That is to say,

$$(v_x, v_y) = \arg\max_{(x,y)\in\Re^2} f(x, y),$$

where $f(x,y)$ is the profit function for the optimization.

For the direct approaches for vanishing point detection, the search space is $\Re^2$. Obviously, search in such a space is computationally expensive.

We propose a novel and hybrid approach for vanishing point detection. Our approach first votes and clusters line intersections into vanishing point candidates (an indirect approach). Then projection analysis from perspective views on these candidates is performed, which is a direct approach. The vanishing point is obtained by an optimization of a function based on the previous two steps. The function can be expressed as following:

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.36

$$g(x,y) = G(f_{indirect}(x,y), f_{direct}(x,y)), \quad (1)$$

where $f_{indirect}(x,y)$ and $f_{direct}(x,y)$ are the profit functions for the indirect and direct approaches respectively.

For vanishing point detection, first, we locate all straight and illusory lines. Then calculate all intersections for every line pair. These points are partitioned into several groups by clustering. The cluster center point is regarded as a typical representation of its sub region, which is

$$C = \{c_i \mid c_i \in S_i, i = 1,...,N\}. \quad (2)$$

Rather than searching on the whole space in $\Re^2$, we search on the representative point set in C for speed up. Since the point set in C is representative enough, the searched maximum in C is a good approximation to the global maximum. Consequently, the final resulting vanishing point is given by

$$(v_x, v_y) = \underset{(x,y)\in C}{\arg\max}\, g(x,y).$$

Then we perform a direct approach, e.g., projection analysis from a perspective view, on the new search space. Compared C in Equation (2) with $\Re^2$, the search space of our hybrid approach is just composed by several points, which is much smaller than that of the direct approaches. Hence, it is time efficient. If the number of detected lines is enough, sufficient line intersections will be generated. And the true vanishing point will be embedded in these intersections with a high probability.

III. ROBUST VANISHING POINT DETECTION

In a horizontal document, each text row has a clue horizontal direction. Our method uses document boundaries and text rows to detect the horizontal vanishing point. However, in the vertical direction of a horizontal document, there will be no text columns for vertical clues. Similar with the method in [7], we extract the vertical character strokes as illusory vertical lines. We perform robust detection of text baselines and character tilt orientations by heuristic rules and statistical analysis [15].

A. Clustering for Vanishing Point Candidates

For vanishing point detection in document analysis, most researchers directly voted vanishing points with intersected lines or other information on the image plane. Firstly, we clustered line intersections to vanishing points on the image plane. As we known, on the image plane, lines and line intersections are more sensitive than the Gaussian sphere when the perspective distortion is weak. In such cases, line intersection points are distributed in a large (even nearly infinite) range on the image. Secondly, we perform vanishing point candidate clustering on the Gaussian sphere.

(1) Clustering on the image plane

All horizontal lines (including straight lines and smeared lines detected) are

$$Set(Line) = \{(a_1, b_1, c_1), ..., (a_N, b_N, c_N)\},$$

where N is the number of all horizontal lines.

As we know, two lines will produce an intersection point. As a result, there are $N_P = N \times (N-1)/2$ intersections which are possible candidates of the horizontal vanishing point. These intersections are

$$Set(Pt) = \{(x_1, y_1), ..., (x_{N_P}, y_{N_P})\}.$$

After checking the distribution of line intersections, we discover that these intersections are located in the *2D* space with one or more groups with high density. It is natural to partition these points into groups by clustering. A sample of the distribution of intersections for a horizontal vanishing point is described in Fig. 1.

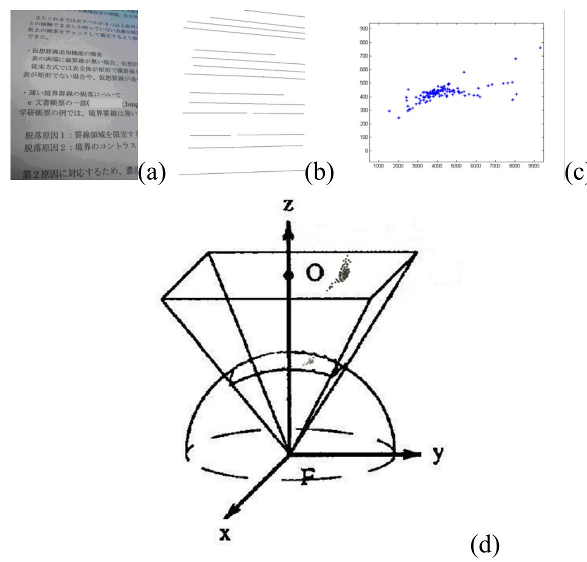

Fig. 1. Intersection distribution for a horizontal vanishing point: (a) captured image, (b) horizontal lines, (c) point distribution on the image plane, and (d) circle intersection distribution on the Gaussian sphere.

Our clustering space is 2D Euclidean space, and the similarity measure of two points is the Euclidean distance,

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2},$$

where (x_i, y_i) is the feature vector (a point in the space).

The k-means clustering algorithm is a rather robust clustering algorithm, which is also proved from our initial experiences. The number of clusters in k-means is specified by the following simple rule:

$$N_{cluster} = \max(\lceil \ln(N_P) \rceil,\ 10).$$

(2) Clustering on the Gaussian sphere

In order to deal with Hough-based problems and the issues of viewpoint sensitivity and image noise for vanishing point detection in perspective images [17-19], alternatively, we propose a clustering based method for detecting vanishing points, which directly clustering projection intersections (circle intersections) from the intersection points of all pairs of lines in the image onto the Gaussian sphere. And we use a rough estimation of the focal length to construct circle intersections on the Gaussian sphere. It is reasonable to cluster these intersections into vanishing points with vector representation on the Gaussian sphere.

Without loss of generality, we also describe the clustering based method for locating the horizontal vanishing point. Each horizontal line on the image plane decides a great circle on the Gaussian sphere. And two lines decide two great circles. These two circles have two intersection points. In fact, only one-half sphere is used, and we get one intersection which is projected from the line intersection point on the image plane [16]. These intersections are possible candidates of the horizontal vanishing point with vector representation on the Gaussian sphere.

After checking the distribution of circle intersections, we find that these intersections are located in the 3D surface with one or more groups with high density. It is natural to partition these points into groups by clustering. A sample of the intersection distribution for a horizontal vanishing point on the image plane is described in Fig. 1 (c). And the corresponding circle intersection distribution on the Gaussian sphere is shown in Fig. 1 (d).

B. Vanishing Point Detection and Selection

After clustering, we will obtain Ncluster clusters, and the center of each cluster is the average of all the intersection points in this cluster. Then, we go back to calculate the vanishing point candidates on the image plane from results on the Gaussian sphere. And the following discussions are based on the image plane.

Suppose Ni is the point number in the ith cluster, each candidate has a weight from clustering which is given by

$$w_i^c = N_i / \sum_{i=1}^{N_{cluster}} N_i.$$

Each weight can be regarded as the profit function for the indirect approach, i.e.,

$$f_{indirect}(x_i, y_i) = w_i^c(x_i, y_i).$$

This weight can be viewed as a conditional probability

$$P((x_i, y_i) \mid A_{indirect}) = f_{indirect}(x_i, y_i),$$

where Aindirect means the vanishing point is detected by the indirect approach.

In order to get a more stable vanishing point, we use a direct approach to refine vanishing point candidates in the above search space [14][15].

For each cluster center, the perspective projection analysis [2] is performed, and the derivative-squared-sum of the projection profiles from a perspective view is calculated, which is normalized as a weight value. Similarly, this weight can also be seen as a conditional probability

$$P((x_i, y_i) \mid A_{direct}) = f_{direct}(x_i, y_i),$$

where Adirect means the vanishing point is derived from the direct approach.

And the combined probability of the above two probabilities is

$$P(x_i, y_i) = \frac{1}{2} f_{indirect}(x_i, y_i) + \frac{1}{2} f_{direct}(x_i, y_i).$$

And the resulting horizontal vanishing point is decided by the Bayesian criteria

$$(v_x, v_y) = \arg\max_{(x_i, y_i) \in X_C} P(x_i, y_i).$$

The last step is to confirm the resulting vanishing point. Our rejection strategy is that the derivative-squared-sum of the true vanishing point will be larger than values of other points [14][15].

IV. EXPERIMENTS

The experiment database includes 418 test samples captured by several mobile phone cameras. These images are in RGB color format with a 1280×960 resolution. More than 90% of the images have perspective distortions.

Given a resulting vanishing point, VP, the relative distance from the ground truth VPt is calculated. If

$$\frac{|VP - VP_t|}{|VP_t|} < T_{VP}, \qquad (3)$$

then VP is regarded as a correct vanishing point. In our system, the ground truth vanishing points are calculated from the manually marked horizontal and vertical lines. When the difference satisfies Equation (3) (TVP=1/20), then there is no seemingly perspective distortion.

We divide our rectified images into five groups: (1) “HIT”, successful for rectification in both horizontal and vertical directions; (2) “HHIT”, successful in the horizontal direction; (3) “VHIT”, successful in the vertical direction; (4) “REJ”, the rectified image is the same as the original image; (5) “ERR”, error rectification.

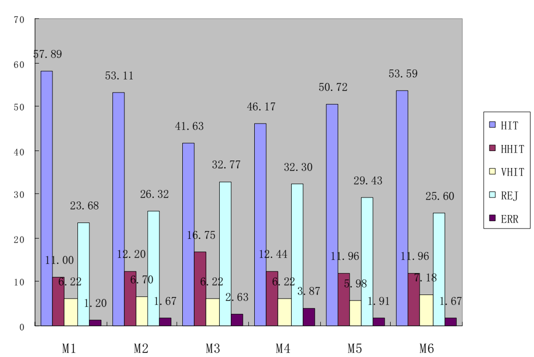

Fig. 2. Rectification accuracy (%) of each resulting group.

Our methods include M1 and M2. M1 is similar to M2 which are both described in this paper. The main difference between them is that M2 clusters vanishing point candidates on the image plane, while M1 performs vanishing point candidate clustering on the Gaussian sphere. We compared our methods to other four methods: M3 doesn’t use character vertical strokes for vertical vanishing point detection; M4 uses the indirect approach based on clustering only to detect vanishing points on the image plane; M5 uses model fitting in [9] instead of clustering; and M6 uses a sequential correction style. The accuracy results are described in Fig. 2.

In our dataset, there are even some street signs and non-document images. The fraction of these non-document images is about 20%. The “REJ” rates of our methods are 23.68% and 26.32%, which is mainly caused by too large distortions. For a mobile phone with some proper interactive GUIs, users may accept the results of HIT, HHIT, VHIT, and REJ because the resulting image from these has a much better quality than (or a same quality as) the originally captured image. In this way, the acceptance rates of our methods are 98.80% and 98.33%.

Compared with M3, our methods (M1 and M2) improve the “HIT” groups by 16.26% and 11.48% respectively. This shows that character vertical strokes are very useful to detect the vertical vanishing point for documents without vertical boundaries. Compared with M4, our methods improve 11.72% and 6.94% for the “HIT” accuracy. Our hybrid approach is more robust for vanishing point detection. Compared with M5, our methods improve the “HIT” accuracy by 7.17% and 2.39% and decrease the processing time by 11ms and 12ms, which shows that our clustering strategy is robust and fast compared to the traditional model fitting. Our methods have a similar performance with M6, but M6 uses a sequential style with partial rectification.

The processing speed is shown in Table 1, where “Time” represents the average processing time for each image without including the time for the grayscale image conversion and the final perspective transformation. Experiments are run on a DELL PC with 3GHz CPU, 2G Memory on a Windows XP OS.

Table 1. Results of the average processing time.

	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Time (ms)	108	103	72	90	115	226

As shown in Table 1, the average processing time of our methods is largely less than M6, and the reduced time is more than 100ms. We also test the direct approach by a hierarchical search for horizontal vanishing point detection in [2], which is more time consuming. For one image in test samples, the detection time is more than one second.

Compared to M2, our new approach (M1) has a higher accurate rate. The “HIT” accuracy is improved from 53.11% to 57.89%. This shows that vanishing point candidate clustering on the Gaussian sphere is effective. Moreover, the additional processing time is only 5ms.

Our rectification technology for perspective document images has been implemented and applied into real mobile phones. Real applications show that our method has an acceptable performance for both accuracy and speed. For a mobile phone camera-based document image (with a 1280*960 resolution), the average processing time is about 1s~2s. Some real samples and corresponding rectified images are shown in Fig. 3.


(a)
(b)

Fig. 3. Some real samples captured by mobiles: (a) original images, (b) rectified images.

V. Conclusions

Perspective rectification of MobileCam-based documents faces several challenges, such as speed, robustness, non-boundary documents, etc. In this paper, we present a fast and robust technology to deal with these problems. In our methods, the hybrid approach for vanishing point detection combines direct and indirect approaches with high precision and fast speed. The experiments on different document images captured by mobile phone cameras show that our method has a good performance with an average speed of about 100ms on a regular PC. Moreover, our perspective rectification system has been applied into real mobile phones.

Acknowledgments

The work of the authors Xu-Cheng Yin and Hong-Wei Hao is partly supported by the R&D Special Fund for Public Welfare Industry (Meteorology) of China under Grant No. GYHY201106039 and the Fundamental Research Funds for the Central Universities under Grant No. FRF-BR-10-034B.

References

[1] J.F. Canny, “A computational approach to edge detection,” *IEEE Trans. on PAMI*, vol. 8, no. 6, pp. 679-698, 1986.

[2] P. Clark, and M. Mirmehdi, “Rectifying perspective views of text in 3D scenes using vanishing points,” *Pattern Recognition*, vol. 36, no. 11, pp. 2673-2686, 2003.

[3] C.R. Dance, “Perspective estimation for document images,” *Proceedings of SPIE Conference on Document Recognition and Retrieval IX*, pp. 244-254, 2002.

[4] W.L. Hwang, C.-S. Lu, and P.-C. Chung, “Shape from texture: estimation of planar surface orientation through the ridge surfaces of continuous Wavelet transform,” *IEEE Trans. on Image Processing*, vol. 7, no. 5, pp. 773-780, 1998.

[5] D.X. Le, G.R. Thoma, and H. Wechsler, “Automated borders detection and adaptive segmentation for binary document images,” *Proceedings of ICPR*, pp. 737-741, 1996.

[6] J. Liang, D. Doermann, and H.P. Li, “Camera-based analysis of text and documents: a survey,” *International Journal on Document Analysis and Recognition*, vol. 7, no. 2-3, pp. 84-104, 2005.

[7] S.J. Lu, B.M. Chen, and C.C. Ko, "Perspective rectification of document images using fuzzy set and morphological operations," *Image and Vision Computing*, vol. 23, no. 5, pp. 541-553, 2005.

[8] C. Monnier, V. Ablavsky, S. Holden, and M. Snorrason, "Sequential correction of perspective warp in camera-based documents," *Proceedings of ICDAR*, vol. 1, pp. 394-398, 2005.

[9] M. Pilu, "Extraction of illusory linear clues in perspectively skewed documents," *Proceedings of CVPR*, pp. 363-368, 2001.

[10] S. Pollard, and M. Pilu, "Building cameras for capturing documents," *International Journal on Document Analysis and Recognition*, vol. 7, no. 2-3, pp. 123-137, 2005.

[11] J. Shufelt, "Performance evaluation and analysis of vanishing point detection techniques," *IEEE Trans. on PAMI*, vol. 21, no. 3, pp. 282-288, 1999.

[12] J. Liang, D. DeMenthon, and D. Doerman, "Flattening curved documents in images," *Proceedings of CVPR*, pp. 338-345, 2005.

[13] J. Liang, D. DeMenthon, and D. Doerman, "Camera-based document image mosaicing," *Proceedings of ICPR*, pp. 476-479, 2006.

[14] X.-C. Yin, J. Sun, S. Naoi, et al., "A multi-stage strategy to perspective rectification for mobile phone camera-based document images," *Proceedings of ICDAR*, pp. 574-578, 2007.

[15] X.-C. Yin, J. Sun, S. Naoi, et al., "Perspective rectification for mobile phone camera-based documents using a hybrid approach to vanishing point detection," *Proceedings of International Workshop on CBDAR*, pp. 37-44, 2007.

[16] S. Barnard, "Interpreting perspective images," *Artificial Intelligence*, vol. 21, no. 4, pp. 435-462, 1983.

[17] R.T. Collins, and R.S. Weiss, "Vanishing point calculation as a statistical inference on the unit sphere," *Proceedings of ICCV*, pp. 400-403, 1990.

[18] A. Minagawa, N. Tagawa, T. Moriya, T. Gotoh, "Vanishing point and vanishing line estimation with line clustering," *IEICE Trans. on Information and Systems*, vol. E83-D, no. 7, pp. 1574-1581, 2000.

[19] E. Lutton, H. Maitre, and J. Lopez-Krahe, "Contribution to the determination of vanishing points using Hough transform," *IEEE Trans. on PAMI*, vol. 16, no. 4, pp. 430-438, 1994.