SHELF DETECTION VIA VANISHING POINT AND RADIAL PROJECTION

Jian Fan, Tong Zhang

Hewlett-Packard Laboratories, Palo Alto, California, USA

ABSTRACT

Retailers need to monitor products on store shelves in order to maintain adequate stocking and in many cases to comply with a placement arrangement defined by a planogram. It is possible that such inspections be performed automatically or semi-automatically using computer vision techniques. Towards this goal, we propose an algorithm to detect shelves from images captured with a handheld digital camera. First, we identify the vanishing point and the associated line segments corresponding to the shelves. Second, we divide the image into equal-angle wedges centered at the vanishing point and project the associated line segments into the wedges. Finally, we identify shelves by analyzing the projections.

Index Terms— visual shelf monitoring, vanishing point detection

1. INTRODUCTION

The monitoring of store shelves is important for efficient utilization of shelf space and for ensuring compliance of shelf layout with a planogram. Currently, shelf monitoring is most often manual, a slow and laborious process that requires store clerks to visually inspect and take note on shelves. An automated or semi-automated solution to this task, as may come from computer vision techniques, is needed. For example, a video camera may be mounted to monitor a shelf and send captured video to a computer for analysis. For a semi-automated system, store clerks may use their cellphones to take snapshots of shelves and send them to a server for automatic processing.

The literature on the described problem is limited. An inventory monitoring system was patented with focus on the system architecture and the integration of existing techniques [1]. On the technology side, Merler et al. created a new database of 120 grocery products and presented the results of applying three object recognition algorithms to the dataset [2]. For real grocery images captured in frontal view, the best-performing algorithm using SIFT features only achieved 72% overall recall and 18% overall precision, indicating unsatisfactory performance using existing techniques.

The existence of shelves is a prominent feature for this application. A picture of shelves in a supermarket is shown in Figure 1a. Detection of shelves helps object segmentation and viewpoint rectification, two important aspects of object recognition. In this paper, we propose a method to detect shelf lines from images captured with hand-held cameras. Our method makes use of both local features of edges and lines as well as the global feature of vanishing point. The main novelty of our method is the radial projection that measures the coverage of the field of view by line segments.

(a) (b)

Figure 1. (a) An example of shelves in a supermarket. (b) Line segments detected from the image. Horizontal and vertical lines are colored green and blue, respectively.

In the next section, we review related research works. In Section 3, the details of the proposed method are described. Experimental results are presented in Section 4. We conclude the paper in Section 5.

2. RELATED WORK

Edge detection has been studied extensively. Canny's method is the classic for grayscale images [3]. To utilize color information, Di Zenzo proposed a color gradient operator [4]. A computationally simpler color gradient operator combines R, G, B channel gradients pixel by pixel [5].

Line segments are one kind of local feature widely adopted by many computer vision algorithms. A classic method for line detection is the Hough transform. Although this method is advantageous in detecting isolated line segments and is robust to noisy background, it may require significant computing resources and may result in false alarms in which accidentally collinear edges are detected. Alternatively, Jang and Hong proposed a line segment grouping method combining local line segment finding and a global accumulator array [6]. In a more recent paper, von Gioi et al. proposed a fast line segment detector (LSD) based on gradient orientations instead of edges [7].

Vanishing points provide important cues in a manmade environment where many parallel lines exist. Among the

vast amount of work on vanishing point detection, three have the most influence on the present work. Magee and Aggarwal proposed a method based on the grouping of line segment intersections on the Gaussian sphere [8]. Rother used the image plane as the accumulator space and intersection points of all pairs of line segments as accumulator cells [9]. Rother's method also defined a distance measure between a vanishing point and a line segment. Using similar accumulator planes and cells, as well as the J-Linkage algorithm, Tardif proposed a non-iterative approach [10].

The problem of road detection bears some similarities with the present work. Since most manmade roads are bounded by two parallel lines, the vanishing point would provide a strong clue for road boundaries. Kong et al. applied Gabor filters for texture orientation estimation and used the result to vote for a vanishing point [11]. However, road detection differs from shelf detection in several key aspects. First, posing a problem for road detection algorithms, road boundaries are not necessarily manifest as sharp edges. In contrast, edges and line segments corresponding to shelves and the products on them are not difficult to detect. However, distinguishing shelf lines from other lines is not straightforward. Second, while two lines are sufficient to determine a road, the number of shelves is unknown.

3. PROPOSED METHOD

The proposed method makes use of several basic assumptions. We assume that an image contains multiple shelves in horizontal orientation and that the scale of shelves is significantly larger than products they hold. We also assume that the images are uncorrected for lens distortions such that shelves captured in an image may be slightly curved.

The proposed method consists of three stages. First, line segments are detected from a downscaled and then Gaussian smoothed image. In the subsequent stage, the vanishing point corresponding to the shelves is identified from horizontal line segments. In the third stage, the image is divided into wedges centered at the vanishing point, and the line segments associated with the vanishing point are projected onto the wedges. Finally the wedges corresponding to shelves are determined by analyzing the projection results.

3.1. Detection of edges and line segments

We adopt a Canny-type edge detector. To take advantage of color information, the grayscale gradient operator was replaced with a color gradient operator $\nabla I_{r,g,b} = \max_{\|x\|}\{\nabla I_r, \nabla I_g, \nabla I_b\}$ [5].

To detect line segments from edges, we adopt an edge-following approach. The edge-following procedure starts from an unvisited edge pixel. From the current edge pixel location, its eight neighbors are checked. A new edge pixel is added to the collection if it satisfies the following conditions: 1) it is at an unvisited location, 2) its gradient orientation is within a set range relative to the existing collection, and 3) the new collection of points still fit to a line with a given criterion. This procedure is repeated until no new edge pixel can be added. Figure 1b shows line segments detected from an image. Note that in general, a shelf level consists of multiple line segments due to lens-induced geometric distortion, disjointed shelves or image noise. Our shelf detection method does not require linking of line segments.

3.2. Detection of the vanishing point of the shelves

We adopt the method proposed by Magee and Aggarwal for the representation of a vanishing point [8]. Assuming the image plane as $Z=1$, two line segments l_1 and l_2 on the image plane are defined as $l_1: \vec{p}_1 = (x_1, y_1, 1) \leftrightarrow \vec{p}_2 = (x_2, y_2, 1)$ and $l_2: \vec{p}_3 = (x_3, y_3, 1) \leftrightarrow \vec{p}_4 = (x_4, y_4, 1)$. The three points $(0,0,0)$, $\vec{p}_1$, and $\vec{p}_2$ form a 3D plane, and likewise, the three points $(0,0,0)$, $\vec{p}_3$ and $\vec{p}_4$ form another 3D plane. Since the intersection of the two planes consists of the vanishing line, it can be shown that the direction vector of the vanishing line is

$$\vec{v} = \frac{(\vec{p}_1 \times \vec{p}_2) \times (\vec{p}_3 \times \vec{p}_4)}{\left|(\vec{p}_1 \times \vec{p}_2) \times (\vec{p}_3 \times \vec{p}_4)\right|},$$

where the symbol $\times$ represents the vector cross-product and $||$ represents the vector norm. If the vanishing line is not parallel to the image plane ($v_z \neq 0$), the vanishing point at the image plane can be computed by $x_v = v_x / v_z$ and $y_v = v_y / v_z$.

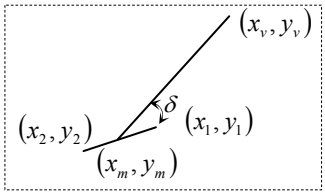

Figure 2. Angle between a line segment and a vanishing point.

For a line segment in the image plane with two end points (x_1, y_1) and (x_2, y_2), we adopt Rother's definition for its closeness to a vanishing point as the angle between the line segment and the line connecting the vanishing point and the midpoint of the segment, as illustrated in Figure 2.

Our vanishing point detection consists of the following steps:

1. Randomly select up to *M* (ex., *M*=500) pairs of horizontal line segments (oriented at less than a 45-degree angle relative to the horizon)
2. Compute the vanishing point (VP) for each pair and use them as accumulator cells
3. For each horizontal line segment, compute its angle with every vanishing point cell constructed in step 2. If the angle is within a given threshold (ex., 3 degrees), we add the line length value to the

corresponding VP cell accumulator, and associate the line segment with the VP.

4. Search all VP cells and find the one with the highest accumulated value as the VP of the shelves.

3.3. Radial projection and shelf detection

Once the vanishing point is identified, we divide the image into equally spaced wedges centered at the vanishing point. A wedge w_i is specified by a center line at angle θ_i and two border lines at angles $(\theta_i - \alpha/2)$ and $(\theta_i + \alpha/2)$, as illustrated in Figure 3a. The center line of a wedge (referred to as wedge lines in the following discussion) shall intersect with image boundaries at two points p_1^i and p_2^i. We call the line interval $\left[p_1^i, p_2^i\right]$ as the intersection of w_i. The length L_i of the intersection varies depending on the angle θ_i of the line. An example of the wedge lines drawn on a real-world image of shelves (Figure 1) is shown in Figure 3b.

The number of wedges should be determined by both the angle spanning the image region from the vanishing point and the minimum angle between two adjacent shelves. According to the sampling theorem, the angle $\Delta\theta$ between two adjacent wedges must not exceed half of the minimum angle between two adjacent shelves. In addition, wedges may or may not overlap, depending on the angle α. The effect of these key parameters will become clear in the experiments section.

A line segment can be projected onto a wedge if the angle between the line and the wedge line is within the range of $\pm\alpha/2$. Figure 3c illustrates a line segment l_1 projected onto the line l_2. The projected interval is denoted as $\overrightarrow{u_0u_1}$ using line l_2 as the new axis.

The projection of multiple lines onto a wedge w_i may be measured in two aspects:

1. Coverage of the intersection. We associate each wedge w_i with an array of integer counters C_i with $\lfloor L_i \rfloor$ elements, where $\lfloor x \rfloor$ stands for the floor function and L_i is the length of the intersection. The counters are initialized to zero. For every line l_j projected to the wedge, the corresponding counters within the indices of $\lfloor u_0 \rfloor$ and $\lfloor u_1 \rfloor$ of C_i are incremented by one. After taking all the projections into account, a zero at $C_i[k]$ indicates a gap at the intersection. Gaps are quantitatively characterized by two metrics:
   1) Normalized sum of all gaps G_{sum}^i. This is the total number of counters with zero divided by the length of the intersection L_i.
   2) Normalized maximum gap $G_{\max}^i$. This is the maximum gap, or the maximum run length of consecutive zeros divided by the length of the intersection L_i. Notice that $G_{\max}^i \le G_{sum}^i$.
2. Closeness of lines to the wedge line. This metric of the weighted sum of projected line lengths is computed as: $S_i = \sum_{j=1}^{N} \frac{l_j \cdot \cos\delta_{j,i}}{\left(1 + d_{j,1} + d_{j,2}\right)}$, where l_j is a line segment, $\delta_{j,i}$ is the angle between the line l_j and the wedge line i, and $d_{j,1}$ and $d_{j,2}$ are the distances between the two endpoints of the line l_j to the wedge line, respectively.

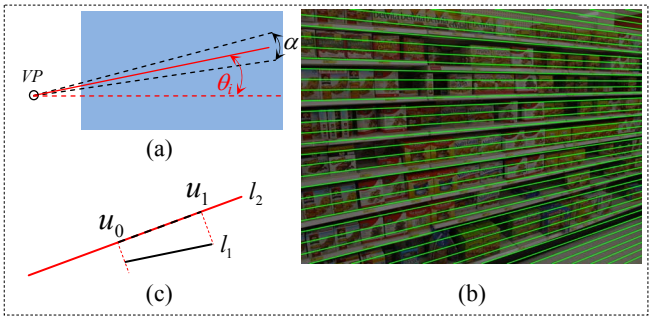

Figure 3. (a) One wedge, (b) wedge lines overlaid on a real image, and (c) a line segment projected onto a wedge.

Shelves are then identified in the following two steps.

1. Identify and flag wedges with near-complete coverage such that $\left(G_{sum}^i < TH_1\right) \wedge \left(G_{max}^i < TH_2\right)$. When wedges overlap or a shelf contains multiple edges, multiple wedges may correspond to one shelf.
2. For a group of consecutive flagged wedges, compute a weighted average angle using the metrics $\{S_i\}$ computed above.

Figure 4 shows an example of the two detection steps. The blue circles are the wedges flagged in the first step and the red lines are the positions computed by the weighted average of corresponding metrics $\{S_i\}$.

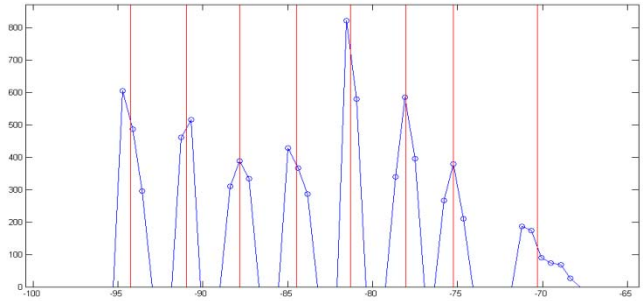

Figure 4. Wedges flagged with high coverage (blue circles) and shelve positions identified (red lines). The horizontal axis is the angles and the vertical axis is the metric *S* of the wedges.

4. EXPERIMENTAL RESULTS

The performance of the proposed method was evaluated using a set of seventeen images captured at local supermarkets and department stores with a Canon G10 camera. Products on the shelves include shoe boxes, toy boxes, bottled drinks, shampoo bottles, canned foods and candy bags. Four of the test images are shown in Figure 5. Notice that some shelves are partially occluded by price tags.

The images were downscaled to 1200x900 and smoothed with a 5x5 Gaussian kernel of $\sigma = 0.8$. Figure 1b shows line segments detected from the image. The two thresholds are set as TH_1= 0.8 and TH_2= 0.2. Two examples of detected shelves are shown in Figure 6. For the result in the left panel of Figure 6, we counted sixe true detections and two false alarms. In the right panel of Figure 6, we counted five true detections and one miss (the very top shelf).

We used visual inspection and manual counting to evaluate performance on various configurations of wedges, as shown in Table 1. For the non-overlapping configurations of 1 and 2, the detection performance is very similar. The low recall is due to missed detections. The algorithm performed substantially better on the configurations 3 and 4 of overlapped wedges. Configuration 4, which had the most overlap, achieved the best performance, with the F-measure reaching 95%. The false alarms mostly resulted from tightly placed boxes, while the missed detections are mainly due to curved lines or disjointed shelves.

Figure 5. Sample test images. (top left) candy bags, (top right) shoe boxes, (bottom left) shampoo bottles with price tags on the shelves, (bottom right) bags and bottles.

Figure 6. Two examples of shelf detection results. Green lines represent detected shelves. See text for more explanation.

Table 1. Wedge parameters and performance.

Config.	$\Delta\theta^o$	α^o	Precision	Recall	F-measure
1	0.992	0.992	0.97	0.71	0.82
2	0.573	0.573	1.00	0.68	0.81
3	0.573	0.802	0.97	0.91	0.94
4	0.573	1.146	0.93	0.98	0.95

5. CONCLUSION

In this paper, we propose a novel approach for shelf detection. Based on line and vanishing point detection, we divide the image region into wedges, introduce a radial projection to measure the spatial coverage of the line segments, and use it as a signature of shelves. Using real-world images, we explored the effect of wedge configuration. Our experiments clearly indicate that overlapped wedges perform better than non-overlapped ones. Our method achieved a best result of 95% in F-measure. To further improve performance, we need to incorporate other contextual image features to distinguish shelves from boxed products.

6. REFERENCES

[1] F. Linaker, R. B. Groenevelt, A. Opalach and A. Fano, "Determination of Inventory Conditions based on Image Processing," *US patent No. 8,009,864*, 2011.

[2] M. Merler, C. Galleguillos and S. Belongie, "Recognizing Groceries in situ Using in vitro Training Data", CVPR, pp. 1-8, 2007.

[3] John Canny, "A computational Approach to Edge Detection", *IEEE Trans. PAMI*, Vol. 8, No. 6, 1986.

[4] Silvano Di Zenzo, "A Note on the Gradient of a Multi-image", *CVGIP*, Vol. 33, pp. 116-125, 1986.

[5] Jian Fan, "A Local Orientation Coherency Weighted Color Gradient for Edge Detection", ICIP 05, pp. 1132-1135, 2005.

[6] Jeong-Hun Jang and Ki-Sang Hong, "Fast Line Segment Grouping Method for Finding Globally More Favorable Line Segments", Pattern Recognition, No. 35, pp.2235–2247, 2002.

[7] R. G. von Gioi, et al, "LSD: a Fast Line Segment Detector with a False Detection Control", IEEE Trans. PAMI, Vol. 32, No. 4, 2010.

[8] M. J. Magee and J. K. Aggarwal, "Determining Vanishing Points from Perspective Images", CVGIP, No. 26, pp. 256-267, 1984.

[9] Carsten Rother, "A New Approach to Vanishing Point Detection in Architectural Environments", Image and Vision Computing, No. 20, pp.647–655, 2002.

[10] Jean-Philippe Tardif, "Non-Iterative Approach for Fast and Accurate Vanishing Point Detection", ICCV, pp. 1250-1257, 2009

[11] Hui Kong, J.-Y. Audibert, and J. Ponce, "Vanishing Point Detection for Road Detection", CVPR, pp. 96–103, 2009.