Computer Vision and Image Understanding 149 (2016) 186–196

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Pedestrian lane detection in unstructured scenes for assistive navigation

Son Lam Phung*, Manh Cuong Le, Abdesselam Bouzerdoum

School of Electrical, Computer, and Telecommunications Engineering, University of Wollongong, Australia

ARTICLE INFO

Article history:
Received 15 April 2015
Accepted 23 January 2016

Keywords:
Pedestrian lane detection
Vanishing point estimation
Assistive and autonomous navigation
Benchmark dataset

ABSTRACT

Automatic detection of the pedestrian lane in a scene is an important task in assistive and autonomous navigation. This paper presents a vision-based algorithm for pedestrian lane detection in unstructured scenes, where lanes vary significantly in color, texture, and shape and are not indicated by any painted markers. In the proposed method, a lane appearance model is constructed adaptively from a sample image region, which is identified automatically from the image vanishing point. This paper also introduces a fast and robust vanishing point estimation method based on the color tensor and dominant orientations of color edge pixels. The proposed pedestrian lane detection method is evaluated on a new benchmark dataset that contains images from various indoor and outdoor scenes with different types of unmarked lanes. Experimental results are presented which demonstrate its efficiency and robustness in comparison with several existing methods.

1. Introduction

Locating the pedestrian lane in a given scene is a key component in many assistive and autonomous navigation systems. It enables a vision-impaired person to find the walkable path and maintain his or her balance while walking – a challenging task that at present is performed mostly using a white cane or a guided dog [1]. It also allows a smart wheelchair to traverse a pedestrian lane with little guidance from the disabled user [2]. Pedestrian lane detection is also useful for autonomous vehicles in sensing off-limit regions or pedestrians in a scene [3]. Furthermore, algorithms for finding the pedestrian lane can be extended to locate open roads for self-driven cars or robots. Pedestrian lane detection in fact complements other features, e.g. obstacle detection [4,5] and GPS-based guidance [6], of electronic navigation devices.

Despite its significance, there are only a few methods proposed for pedestrian lane detection, which are mostly concerned with pedestrian lanes having white markers [7–10]. To address this gap, this paper focuses on camera-based detection of pedestrian lanes in unstructured environments. In this paper, a pedestrian lane is assumed to exist in the scene. However, the scene is under varying lighting conditions and could be indoor or outdoor. Furthermore, the pedestrian lanes can have arbitrary surfaces with no painted markers.

Existing algorithms for unmarked lane detection (including pedestrian lanes) rely on color and texture of lane surfaces to differentiate the lane pixels from the background [11–14]. These algorithms require off-line training, and hence their detection accuracy decreases when the lane appearance differs from the training data. In practice, the lane appearance varies significantly due to different lane surfaces or illumination conditions. Other existing algorithms locate the lane boundaries among the edges that point to the vanishing point of the image [15,16]. However, algorithms based on finding the lane borders are sensitive to background clutter. In this paper, we propose a new method to detect unmarked pedestrian lanes using both color, edge, and shape features. In contrast to the existing methods, the proposed approach constructs a lane model from the input image, and is therefore more adaptive to different illumination conditions and lane surfaces. The main contributions of the paper can be briefly described as follows:

- Firstly, we propose an improved vanishing point estimation method using local orientations of color edge pixels. Estimating the vanishing point using edge pixels is more efficient than using all pixels as done in the existing methods [15,16]. In addition, to estimate local orientations and edge pixels more robustly, we apply the color tensor on multiple color channels, instead of relying on only the intensity channel.
- Secondly, we present a method to define automatically a sample region, from which a lane appearance model is constructed adaptively for each input image. This sample region is determined using the vanishing point and the geometric and color

* Corresponding author. Tel. +61 242213407.
E-mail addresses: phung@uow.edu.au (S.L. Phung), clm635@uowmail.edu.au (M.C. Le), bouzer@uow.edu.au (A. Bouzerdoum).

http://dx.doi.org/10.1016/j.cviu.2016.01.011

features of lane borders and surfaces. The lane model is therefore adaptive to various types of lane surfaces. To make the lane model more robust to the lighting conditions, this paper employs the illumination-invariant color space (IIS). In addition, we propose a novel lane score that combines color, edge, and shape features for detecting unmarked pedestrian lanes.

- Lastly, we create a new dataset with manually annotated ground-truth for objective evaluation of algorithms for vanishing-point estimation and pedestrian-lane detection. Although several datasets exist for road/lane detection for vehicles, our dataset, to the best of our knowledge, is the first for pedestrian lanes. This dataset, collected from realistic indoor and outdoor scenes, with various shapes, textures, and surface colors, is expected to facilitate research in vanishing-point estimation and pedestrian-lane detection. It is available at www.uow.edu.au/~phung/plvp_dataset.html.

The remainder of the paper is organized as follows. Existing methods for lane detection in unstructured scenes are reviewed in Section 2. The proposed method is presented in Section 3. Experimental results and analysis are described in Section 4. Finally, conclusions are given in Section 5.

2. Related work

Current vision-based approaches for detecting pedestrian lanes in unstructured scenes can be divided into two categories: (i) lane segmentation, and (ii) lane-border detection. In the *lane segmentation* approach, off-line color models are used to differentiate the lane pixels from the background [11,12,17,18]. Different color spaces and classifiers have been used. For example, Crisman and Thorpe use Gaussian models in the red–green–blue (RGB) color space to represent the on-road and off-road classes [11]. Also using the RGB space, Tan et al. capture the variability of the road surface with multiple color histograms, and the background with a single color histogram [12]. Instead of using the RGB space, Ramstrom and Christensen employ UV, normalized red and green, and luminance components and construct Gaussian mixture models for the road-surface and background classes [18]. Sotelo et al. employ the hue-saturation-intensity (HSI) color space [17]. In their method, achromatic pixels (i.e. with extreme intensities or low saturations) are classified using intensity only, whereas other pixels are classified by thresholding their chromatic distance to the training colors. Because the color models are trained off-line, these methods do not cope well with the appearance variations in lane surfaces.

To address this problem, several methods model the lane pixels directly from sample regions in the input image [19–22]. These methods determine the sample lane regions in different ways. For example, Alvarez et al. select small random areas at the bottom and in the middle of the input image [22,23]. Miksik et al. initialize the sample lane region as a trapezoid at the bottom and center of the image, and then refine the sample region using the vanishing point [21]. He et al. form a sample lane region from the candidate lane boundaries, which are detected using the vanishing point and an assumption about the lane width [19]. The performance of these methods depends on the quality of the sample regions, which in turn relies on prior knowledge about the walking lane.

In the *lane-border detection* approach, the lane boundaries are determined using the vanishing point [15,16] or templates of the lane boundaries [24]. In [15], the lane borders are detected among the edges pointing to the vanishing point. The optimal left and right edges are judged using an objective function that measures the color and texture differences between lane and non-lane regions. This method is effective only when the lane region is homogeneous and differs significantly from non-lane regions in terms of color and texture. Kong et al. also find the lane borders from the edges pointing to the vanishing point, except that their method ranks edges using texture orientation and color features [16]. Because this method relies only on edges for lane-border detection, it is sensitive to background edges. In another method, the lane boundaries are found from the edges of homogeneous color regions by matching with lane templates [24]. Recently, Chang et al. propose combining lane-border detection and road segmentation for detecting lanes [25]. Similarly to [16], their method detects lane borders using the vanishing point. The lane region is segmented using the color model learned from a homogeneous region at the bottom and middle of the input image. In [26], the two left and right borders of the lane are found among the rays that point to the vanishing point; this approach is suitable mainly for pedestrian lanes with straight-line borders. This paper extends this approach to detect pedestrian lanes with curved borders and varied surfaces.

3. Proposed pedestrian lane detection method

In this section, we present the new method for detecting unstructured pedestrian lanes, which comprises three main stages: (i) vanishing point estimation, (ii) sample region selection, and (iii) lane segmentation.

3.1. Vanishing point estimation

The vanishing point in an image is often located using either line segments [27–29] or local orientations [15,16,30]. For unstructured scenes with non-straight edges, using local orientations is more suitable than using line segments for vanishing point estimation. However, most existing methods based on local orientations have high computational complexity and are sensitive to background clutter. Furthermore, they rely on the intensity channel only, even though color channels provide photometric information that can lead to more robust detection of edges and local orientations. In this paper, we propose to improve the accuracy and efficiency of vanishing point detection, by employing color tensor to capture image structure and focusing on edge pixels only.

The color tensor is a tool for analyzing the local differential structure of a color image [31]. Consider an image with three color channels: $F = \{F_k; k = 1, 2, 3\}$. Let $D_{k,x}$ and $D_{k,y}$ denote the derivatives of F_k along the horizontal and vertical direction, respectively. Let $\mathbf{w}$ be the convolution kernel of a smoothing filter. The color tensor of the image is represented as

$$\begin{pmatrix} G_{xx} & G_{xy} \\ G_{yx} & G_{yy} \end{pmatrix} \quad \text{where} \quad \begin{cases} G_{xx} &= \mathbf{w} * \left[\sum_{k=1}^{3} D_{k,x} \circ D_{k,x} \right] \\ G_{yy} &= \mathbf{w} * \left[\sum_{k=1}^{3} D_{k,y} \circ D_{k,y} \right] \\ G_{xy} &= \mathbf{w} * \left[\sum_{k=1}^{3} D_{k,x} \circ D_{k,y} \right] \end{cases}. \tag{1}$$

Here, $*$ denotes the 2-D convolution, and $\circ$ denotes the element-wise multiplication (Hadamard product). Based on eigenvalue analysis of the color tensor [31], we estimate the dominant local orientation $\boldsymbol{\theta}$ and the edge strength $\boldsymbol{\lambda}$ for all image pixels as

$$\boldsymbol{\theta} = \frac{1}{2} \arctan\left(\frac{2G_{xy}}{G_{xx} - G_{yy}} \right) + \frac{\pi}{2}, \tag{2}$$

$$\boldsymbol{\lambda} = \frac{1}{2}\left(G_{xx} + G_{yy} + \sqrt{(G_{xx} - G_{yy})^2 + 4G_{xy}^2} \right), \tag{3}$$

where the arithmetic operations are performed element-wise. Next, the edge pixels in the image are identified via non-maximum suppression and hysteresis thresholding, as done in the intensity-based Canny edge detector. The main difference in this paper is

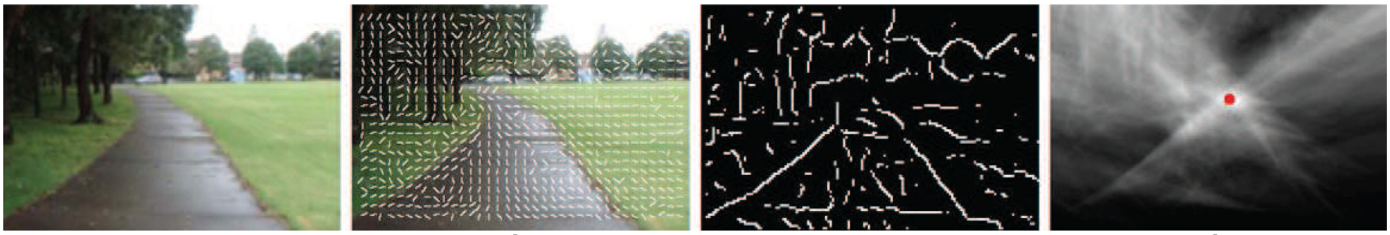

Fig. 1. Illustration of the proposed vanishing point estimation: (a) color input image; (b) local orientations estimated by the color tensor for sampled pixels; (c) edge map obtained by the *color* Canny edge detector; (d) VP map and the vanishing point (in red). See the electronic color image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

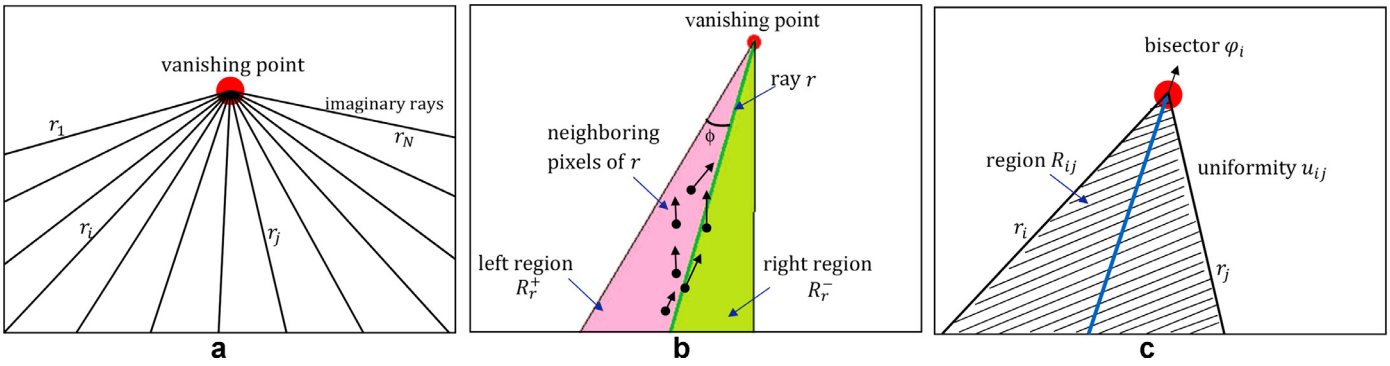

Fig. 2. Selecting the sample lane region: (a) rays created from the vanishing point; (b) properties of a single ray; (c) properties of a pair of rays. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

that the dominant local orientation $\boldsymbol{\theta}$ and the edge strength $\boldsymbol{\lambda}$ are estimated more reliably using photometric information obtained from (2) and (3). For the input image of Fig. 1(a), Fig. 1(b) shows the computed local orientations, and Fig. 1(c) shows the estimated edge map.

To determine the vanishing point (VP), each pixel location $v = (x_v, y_v)$ is considered as a candidate, for which a VP score is computed. Let P be the set of edge pixels $\{p = (x_p, y_p)\}$ where $y_p > y_v$. Let Δ_{vp} be the difference between the dominant local orientation at pixel p and the angle of vector $\boldsymbol{\ell}_{vp}$ connecting v to p: $\Delta_{vp} = |\theta_p - \angle\boldsymbol{\ell}_{vp}|$. Let μ_{vp} be the ratio between the length of $\boldsymbol{\ell}_{vp}$ and the diagonal length L of the image: $\mu_{vp} = |\ell_{vp}|/L$. After investigating several choices, we propose to define the contribution of pixel p to the score of candidate pixel v as

$$s(v, p) = \begin{cases} \exp\left(-\Delta_{vp}\ \mu_{vp}\right), & \text{if } \Delta_{vp} \le \tau_{\mathrm{o}}, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Here, τ_{o} is a positive threshold on the orientation similarity between p and $\boldsymbol{\ell}_{vp}$. Eq. (4) means that $s(v, p)$ is high if (i) pixel p has a similar orientation to vector $\boldsymbol{\ell}_{vp}$, and (ii) pixel p is spatially close to v. The VP score of candidate v is the sum of contributions from all pixels in P:

$$S(v) = \sum_{p \in P} s(v, p). \tag{5}$$

The vanishing point is finally found as the pixel with the highest VP score. Fig. 1(d) illustrates the VP map and the vanishing point computed for the image in Fig. 1(a). More results of the proposed method for vanishing point estimation are given in Section 4.3.

3.2. Sample region selection

Because the appearance (e.g. color, edge, shape, texture) of pedestrian lanes in unstructured scenes varies significantly and is strongly affected by illumination conditions, it is difficult to obtain a robust appearance model with off-line training. Hence, it is more plausible to construct an appearance model adaptively and directly from the input image. To this end, existing methods (e.g. [20,21,23]) usually select the sample region as a small region at the bottom or center of the input image. However, the sample region selected in such a manner tends to include non-lane regions. In our approach, the sample region is automatically defined using the vanishing point (estimated in the previous stage), and then verified using color and orientation features of both lane borders and lane regions.

Although a lane may have various shapes, its main part can be approximated with straight borders. Hence, it is possible to represent the border of the sample region using imaginary rays. To this end, a set of rays $\mathcal{B} = \{r_1, r_2, \ldots, r_N\}$ emanating from the vanishing point is created as shown in Fig. 2(a). These rays are uniformly spaced over an angle range $[\phi_{\min}, \phi_{\max}]$ relative to the horizontal direction. The sample region is identified by finding a ray pair (r_i, r_j) that best captures the main part of the pedestrian lane.

For a given ray r two features are defined: (1) the orientation difference d_{o} between ray r and its neighboring pixels; (2) the color difference d_{c} between two regions adjacent to r, see Fig. 2(b). Let θ_r denote the angle of ray r. Let $\mathcal{N}_r$ be the set of pixels whose Euclidean distance to r is smaller than $L\tau_{\mathrm{e}}$. Here, L is the diagonal length of the image, and τ_{e} is a threshold. The orientation difference d_{o} between r and its neighboring pixels is calculated as

$$d_{\mathrm{o}} = \frac{1}{|\mathcal{N}_r|} \sum_{p \in \mathcal{N}_r} |\theta_r - \theta_p|, \tag{6}$$

where θ_p is the orientation of pixel p computed in (2).

Let R_r^+ and R_r^- be two neighboring regions on the left and right of ray r as shown in Fig. 2(b). These regions are formed from ray r by an angular spacing of ϕ. Suppose that $\mathbf{c}^+$ and $\mathbf{c}^-$ are the mean color of all pixels in R_r^+ and R_r^-, respectively. The color difference

d_c between adjacent regions of ray r is computed as

$$d_c = \frac{||\mathbf{c}^+ - \mathbf{c}^-||_2}{\max(||\mathbf{c}^+||_2, ||\mathbf{c}^-||_2)}, \tag{7}$$

where $||\cdot||_2$ denotes the L_2-norm.

Next, for a given ray pair (r_i, r_j) two features are defined: (1) the color uniformity u_{ij} of pixels between r_i and r_j; (2) the angle ϕ_{ij} of the bisector between r_i and r_j. Let R_{ij} denote an image region formed by a ray pair (r_i, r_j) as shown in Fig. 2(c). The uniformity u_{ij} of R_{ij} is computed as

$$u_{ij} = \sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{k=1}^{M} h(m,n,k)^2, \tag{8}$$

where h is the normalized 3-D color histogram of pixels in R_{ij}, and M is the number of bins for each color channel.

In summary, for a given ray pair (r_i, r_j), six features are extracted: (1) the orientation difference $d_{o,i}$ of ray r_i and its neighboring pixels; (2) the orientation difference $d_{o,j}$ of ray r_j and its neighboring pixels; (3) the color difference $d_{c,i}$ between adjacent regions of ray r_i; (4) the color difference $d_{c,j}$ between adjacent regions of ray r_j; (5) the color uniformity of u_{ij} of region R_{ij}; 6) the bisector angle ϕ_{ij} of ray r_i and r_j.

Given these six features, we propose the following lane score for the ray pair (r_i, r_j):

$$f(r_i, r_j) = f_1(d_{o,i})\, f_1(d_{o,j})\, f_2(d_{c,i})\, f_2(d_{c,j})\, f_3(u_{ij})\, f_4(\phi_{ij}), \tag{9}$$

where the individual score functions are given by:

$$f_1(d_o) = \exp\{-d_o/\pi\}, \tag{10}$$

$$f_2(d_c) = \frac{1}{1 + a\, e^{-b\, d_c}}, \tag{11}$$

$$f_3(u) = \frac{1}{1 + \alpha\, e^{-\beta\, u}}, \tag{12}$$

$$f_4(\phi) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\phi - \overline{\phi})^2}{2\sigma^2}\right\}. \tag{13}$$

In Eqs. (10)–(13), a, b, α, β, σ and $\overline{\phi}$ are fixed parameters that are determined empirically from the training data. The individual score functions are chosen to model the relationship between a feature and the lane score. For example, Eq. (10) means that the smaller is the orientation difference d_o (i.e. when neighboring pixels have similar orientations as ray r), the higher is the score $f_1(d_o)$, and vice versa. Eq. (11) indicates that the higher is the color difference d_c (i.e. when ray r is at the lane border), the higher is the score $f_2(d_c)$. Eq. (12) means that the higher is the color uniformity u, the higher is the score $f_3(u)$. Lastly, Eq. (13) is based on the observation that the bisector angle on training data approximates a normal distribution.

The optimal pair (r_i^*, r_j^*) for the sample region is obtained by maximizing the lane score:

$$(r_i^*, r_j^*) = \underset{(r_i, r_j)\in\mathcal{B}^2}{\arg\max}\, f(r_i, r_j). \tag{14}$$

Fig. 3(a) shows an example of detecting the borders of the sample region. Instead of using the entire triangular region defined by (r_i^*, r_j^*), we use only the trapezoidal region (lower-half) formed by the two rays as a lane sample region, see Fig. 3(b). This strategy is adopted to improve the stability of the lane model, even if the vanishing point is located outside the image or the pedestrian lane.

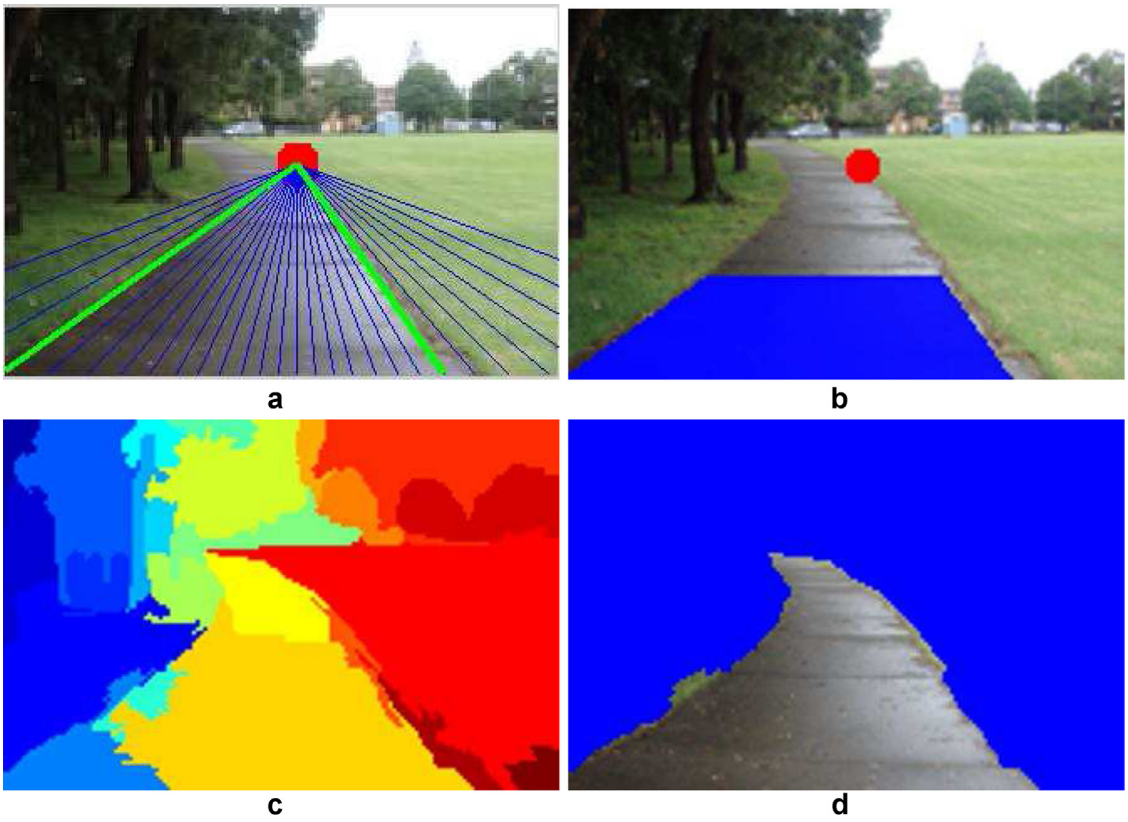

Fig. 3. Illustration of the proposed method for pedestrian lane detection: (a) the imaginary rays (blue lines) and the detected borders (green lines) of the sample region; (b) lane sample region (blue region); (c) color homogeneous sub-regions segmented using the graph-based method [32]; (d) segmented walking lane. See the electronic color image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Lane segmentation

In this stage, the input image is segmented initially into color homogeneous sub-regions. Numerous image segmentation algorithms can be applied. In this paper we use the graph-based segmentation algorithm presented in [32], because it is fast and suitable for our task . This algorithm initializes sub-regions as single pixels. Adjacent sub-regions are then merged iteratively, according to the color difference between the sub-regions. Fig. 3(c) illustrates the segmented regions for the input image of Fig. 1(a).

Next, the pedestrian lane is detected. Let $\mathcal{R} = \{R_1, R_2, \ldots\}$ be the set of color homogeneous sub-regions. The pedestrian lane is treated as a set Z of connected sub-regions of $\mathcal{R}$. Two sub-regions R_i and R_j are considered to be connected if there exist two pixels $p_i \in R_i$ and $p_j \in R_j$ that are connected (e.g. 4-connected pixels).

A connected region $Z \subset \mathcal{R}$ is represented by a color feature and a shape feature. The color feature $\mathbf{c}$ is the mean of all color pixels in Z. The lane score for a given color feature $\mathbf{c}$ is defined as

$$g_1(\mathbf{c}) = p(\mathbf{c}|\mathcal{L}), \tag{15}$$

where $p(\mathbf{c}|\mathcal{L})$ is the class-conditional probability density function for the lane class. It is estimated from the color histogram of all pixels in the sample lane region, which is found as in Section 3.2. In this paper, we consider two color spaces: the red–green–blue (RGB) and the illumination invariant space (IIS). Compared to the RGB, the IIS is less sensitive to illumination conditions and shading. Conversion from the RGB to the IIS is as follows [33]:

$$\begin{cases} C_1 &= \arctan\{R/\max(G,B)\}, \\ C_2 &= \arctan\{G/\max(R,B)\}, \\ C_3 &= \arctan\{B/\max(R,G)\}. \end{cases} \tag{16}$$

The shape feature $\mathbf{s}$ is extracted using the shape contexts proposed in [34]. The shape contexts are known for their robustness to local deformation and partial occlusion, and their invariance to scale and rotation. Consider a shape with sampling points on its contour. The shape context of a sampling point p is the histogram h_p of the angles and distances from the remaining sampling points to p.

The dissimilarity between the shape contexts of two points p and q is represented as

$$C(p,q) = \frac{1}{2}\sum_{k=1}^{K}\frac{[h_p(k) - h_q(k)]^2}{h_p(k) + h_q(k)}, \tag{17}$$

where K is the number of bins of each shape context. On a single shape, the shape contexts of the points p and q are different, i.e. $C(p, q)$ is high. However, on two similar shapes, the shape contexts of two corresponding points p and q are similar, i.e. $C(p, q)$ is low.

Let $\mathcal{T} = \{T_1, T_2, \ldots\}$ be a set of shape templates for pedestrian lanes. Examples of the shape templates obtained from the training data are shown in Fig. 4. To obtain shape feature $\mathbf{s}$, the outer contour of region Z is sampled in a similar way as the templates. The matching cost $D(\mathbf{s}, T)$ between $\mathbf{s}$ and a template T is modeled as

$$D(\mathbf{s},T) = \frac{1}{|\mathbf{s}|}\sum_{p\in\mathbf{s}}\min_{q\in T} C(p,q), \tag{18}$$

where $|\mathbf{s}|$ denotes the number of sampling points on $\mathbf{s}$. The smaller is the matching cost $D(\mathbf{s}, T)$, the higher is the similarity between $\mathbf{s}$ and T. Consequently, the lane score for shape feature $\mathbf{s}$ is defined as

$$g_2(\mathbf{s}) = \exp\left[-\lambda \min_{T\in\mathcal{T}} D(\mathbf{s},T)\right], \tag{19}$$

where λ is a positive scalar determined from the training data.

Collectively, the lane score for region Z with color feature $\mathbf{c}$ and shape feature $\mathbf{s}$ is calculated as

$$g(Z) = g_1(\mathbf{c})\, g_2(\mathbf{s}). \tag{20}$$

The optimal region Z^* of $\mathcal{R}$ is found by maximizing the lane score:

$$Z^* = \arg\max_{Z\subset\mathcal{R}} g(Z). \tag{21}$$

The optimal region Z^* can be obtained with a computational complexity of $O(2^{|\mathcal{R}|})$ via an exhaustive search among the subsets of $\mathcal{R}$. To reduce the computational load, we adopt a greedy-search algorithm [35], which generates Z^* by iteratively adding and removing sub-regions (see Algorithm 1). At each iteration, a sub-region is added to or removed from Z^*, only if the connectivity of the new Z^* is satisfied and the lane score $g(Z^*)$ is increased. In addition, for faster search we consider only sub-regions $R_i \in \mathcal{R}$ with $p(\mathbf{c}|\mathcal{L})$ greater than or equal to a predefined threshold τ_c. Because the number of sub-regions is finite and the operators in Algorithm 1 are deterministic, the algorithm will converge.

Finally, the optimal region Z^* obtained using Algorithm 1 is considered as a pedestrian lane region if

$$g(Z^*) \geq \tau_v, \tag{22}$$

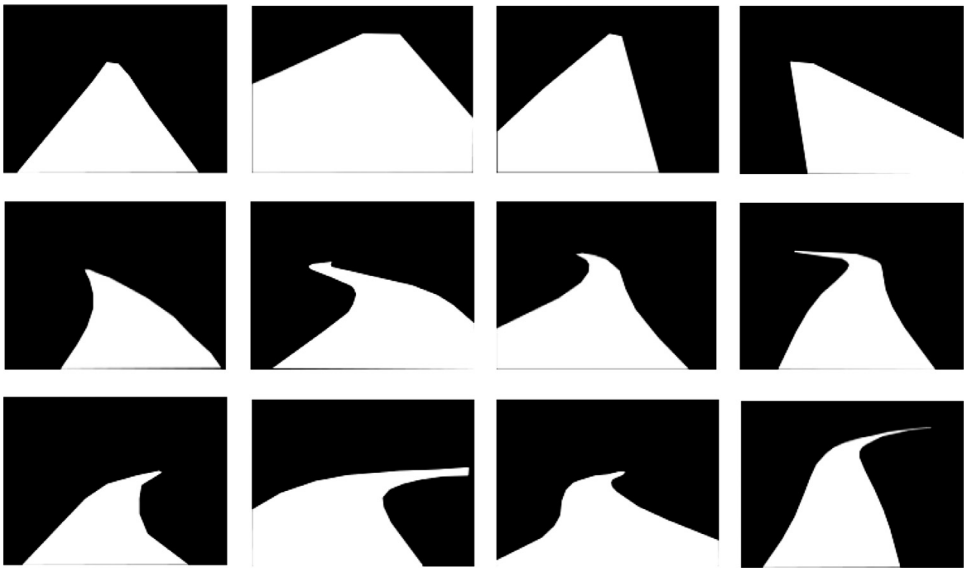

Fig. 4. Example shape templates for pedestrian lanes. Row 1: straight lanes. Row 2: left-curved lanes. Row 3: right-curved lanes.

Algorithm 1 Adding and removing regions for pedestrian lane detection.

$\mathcal{R}' \leftarrow \{R_i \in \mathcal{R} \mid p(\mathbf{c}_i|\mathcal{L}) \geq \tau_c\}$
$Z^* \leftarrow \arg\max_{R_i \in \mathcal{R}'} p(\mathbf{c}_i|\mathcal{L})$
continue $\leftarrow$ TRUE
while continue **do**
  $\mathcal{R}_{\text{add}} \leftarrow \{R_i \in \{\mathcal{R}' - Z^*\}$ so that $Z^* \cup R_i$ is a connected set$\}$
  $R^+ \leftarrow \arg\max_{R_i \in \mathcal{R}_{\text{add}}} g(Z^* \cup R_i)$
  $\mathcal{R}_{\text{rmv}} \leftarrow \{R_i \in Z^*$ so that $\{Z^* - R_i\}$ is a connected set$\}$
  $R^- \leftarrow \arg\max_{R_i \in \mathcal{R}_{\text{rmv}}} g(Z^* - R_i)$
  if $g(Z^* \cup R^+) > g(Z^*)$ **and** $g(Z^* \cup R^+) \geq g(Z^* - R^-)$ **then**
    $Z^* \leftarrow Z^* \cup R^+$
  else if $g(Z^* - R^-) > g(Z^*)$ **then**
    $Z^* \leftarrow Z^* - R^-$
  else
    continue $\leftarrow$ FALSE
  end if
end while

where τ_v is a verification threshold learnt using training data. This step is necessary because the scene may contain no pedestrian lane. Fig. 3(d) illustrates the result of lane detection for the input image shown in Fig. 1(a).

4. Experimental results

In this section, we first describe the image data and evaluation measures (Section 4.1). We then discuss the parameters used in the proposed method (Section 4.2), and present the experimental results of vanishing point estimation (Section 4.3) and pedestrian lane detection (Section 4.4).

4.1. Image dataset and performance measures

We created an image dataset for pedestrian lane detection and vanishing point estimation (PLVP). The dataset consists of 2000 images that were taken under various environmental conditions (indoor and outdoor scenes, different times of day, and different weather conditions). The images contain unmarked pedestrian lanes with various surface structures (pavement, brick, concrete, or soil) and shapes (straight or curved). In many cases, lane regions are affected by extreme lighting conditions (e.g. very low illumination, very high illumination, or strong shadow). To enable quantitative performance evaluation, we manually annotated the lane region and the vanishing point in each image. An example from the PLVP dataset is shown in Fig. 5. Statistics regarding the lane surfaces and the lighting conditions are given in Table 1. The dataset is available for download from www.uow.edu.au/~phung/plvp_dataset.html.

Table 1
Statistics of the PLVP dataset.

Description	Number of images	Percentage (%)
Brick surfaces	637	31.85
Concrete surfaces	944	47.20
Pavement surfaces	179	8.95
Indoor surfaces	159	7.95
Other surfaces	81	4.05
Normal lighting	1393	69.65
Shadows, extreme lighting	607	30.35

To evaluate pedestrian lane detection, the detected regions are compared with the annotated regions. Suppose that R_d is a machine-detected region and R_g is a ground-truth region. The matching score between R_d and R_g is computed as

$$\chi(R_g, R_d) = \frac{|R_g \cap R_d|}{|R_g \cup R_d|}, \tag{23}$$

where $|R|$ denotes the area of region R, $\cap$ denotes the intersection, and $\cup$ denotes the union of R_d and R_g. Detected region R_d is considered as *correct* if there exists a ground-truth region R_g where $\chi(R_g, R_d)$ is greater than or equal to an evaluation threshold τ_e. Similar to the evaluation of other object-detection systems [36,37], τ_e is set to 0.5.

For pedestrian lane detection, three evaluation measures are computed: recall, precision and F-measure. *Recall* is the percentage of the ground-truth lanes that are detected correctly. *Precision* is the percentage of the machine-detected lanes that are considered to be correct. *F-measure* is the harmonic mean of precision and recall:

$$F\text{-}measure = 2 \times \frac{Recall \times Precision}{Recall + Precision}. \tag{24}$$

To evaluate vanishing point estimation, the detected vanishing point is compared with the ground-truth vanishing point. Suppose that P_d is a machine-detected vanishing point, and P_g is the ground-truth vanishing point. Consistently with [30], the estimation error for an image is measured as the ratio of the Euclidean distance from P_d to P_g versus the diagonal length L of the image:

$$E_{\text{vp}} = \frac{|P_d - P_g|}{L}. \tag{25}$$

The averaged estimation error across all test images is used to compare vanishing point estimation algorithms.

4.2. Algorithm parameters

In our experiments, 500 images (image number 1 to 500) were used for training, and 1500 images (image number 501–2000) were used for testing. Note that in the PLVP dataset, images collected from multiple sources were given randomized image

Fig. 5. An example from the PLVP dataset. *Left*: an input color image. *Middle*: the ground-truth pedestrian lane. *Right*: the ground-truth vanishing point. See the electronic color image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Vanishing point estimation performance of the proposed method on the training set for different values of H.

Gaussian window size H	5	7	11	13	17	19
Average error	0.0969	0.0838	0.0737	0.0694	0.0662	0.0783

Table 3
Vanishing point estimation performance of the proposed method on the training set for different values of τ_o.

Angle interval τ_o	$\pi/180$	$\pi/36$	$\pi/18$	$\pi/12$	$\pi/9$
Average error	0.0914	0.0675	0.0659	0.0674	0.0677

numbers. The parameters of the proposed method were selected by analyzing the performance of the pedestrian lane detection on the training set.

For the processing steps described in Section 3.1, two parameters need to be determined: the window size H of the Gaussian filter $\mathbf{w}$, and the angle interval τ_o, see (4). Table 2 shows the VPE error on the training set for different values of H. Based on this table and in order to reduce the filtering time, we selected $H = 13$. Table 3 shows the VPE error on the training set for different values of τ_o. Based on this table and in order to reduce the number of non-zero votes (by using a smaller τ_o), we selected $\tau_o = \pi/36$.

For the steps described in Section 3.2, a similar strategy was adopted to determine the values of the parameters. The number of imaginary rays was selected as $N = 29$, and the angle range of imaginary rays was selected as $[\phi_{\min}, \phi_{\max}] = [\pi/9, 8\pi/9]$. The angular spacing was set as $\phi = \pi/12$. The parameters a and b in (11), α and β in (12), σ and $\overline{\phi}$ in (12) were set as $a = 0.9$, $b = 2.3$, $\alpha = 0.9$, $\beta = 2.3$, and $\sigma = 0.3437$ and $\overline{\phi} = 1.5970$.

For the steps described in Section 3.3, the thresholds τ_c and τ_v were selected as $\tau_c = 0.02$ and $\tau_v = 0.01$. Table 4 shows the lane detection performance of the proposed method on the training set for different values of the shape parameter λ in (19). Based on this result, we selected $\lambda = 5.0$. Table 5 shows the lane detection performance of the proposed method on the training set for different numbers of color quantization bins M. The F-measure fluctuated slightly for $M = 16, 32, 64$, and reduced when $M = 128, 256$. Based on this result, we selected $M = 16$ for our experiments.

4.3. Analysis of vanishing point estimation

The proposed method for vanishing point estimation was compared with three existing methods.

- *Hough-based method* [27]: This method first applies the Hough transform on the edge map to find line segments. It then computes the vanishing point by voting the intersections of line pairs in another Hough transform. In the experiments, we used the same edge map as in the proposed method. The distance and orientation resolutions in the Hough transforms were tuned using the training set.
- *Gabor-based method* [16]: This method applies Gabor filters on the intensity image to compute local orientations, and then estimates the vanishing point using these orientations. Each pixel location v in the top 90% region of the image is considered as a VP candidate. It is voted by all pixels p in the half-disk region, which is centered on v and below v. Our experiments used the MATLAB code provided by the authors of [16]. However, the parameters of the Gabor-based method were tuned using the training set.
- *Optimal local dominant orientation method* (OLDOM) [30]: This method uses four Gabor filters to estimate the local dominant orientation θ at each pixel p in the intensity image. The upward ray r originating from pixel p and along orientation θ is then identified. Each pixel along ray r will accumulate a voting score according to its distance to pixel p. Finally, the image pixel with the highest voting score is considered as the image vanishing point.

Table 5
Lane detection performance of the proposed method on the training set for different color bin numbers.

Number of color bins M	16	32	64	128	256
Recall (%)	93.0	92.8	92.6	92.2	92.2
Precision (%)	97.5	97.5	97.5	97.1	97.1
F-measure (%)	95.2	95.1	95.0	94.6	94.6

Table 6
Performance of vanishing point estimation algorithms on the test set.

Method	Average error	Computation times
Hough-based method (Wang et al. [27])	0.1199 ± 0.1802	0.029
Gabor-based method (Kong et al. [16])	0.0809 ± 0.1034	2.980
OLDOM (Moghadam et al. [30])	0.1964 ± 0.1086	0.595
Proposed VPE method	**0.0707 ± 0.0954**	**0.595**

Table 6 shows the performance of different VPE algorithms on the test set of 1500 images. The average error of the proposed method (0.0707) was significantly lower than that of the Hough-based method (0.1199) and the OLDOM (0.1964). The Hough-based method employs straight lines for finding the vanishing point. It does not work well for natural scenes that contain many non-straight edges. The OLDOM is designed for speed and it uses only four Gabor filters to estimate the local dominant orientations [30]. Furthermore, to calculate the voting score, the OLDOM uses only pixel distance, whereas our method takes into account both pixel distance and pixel orientation difference. The proposed method also had a lower average error (0.0707) compared to the Gabor-based method (0.0809) [16]. The Gabor-based method calculates the voting score for each vanishing point candidate from all pixels in a half-disk region, and is therefore affected by clutter pixels. Furthermore, the Gabor-based method uses only intensity for computing the edge orientations and magnitudes. In comparison, the proposed method uses only edge pixels for voting, and therefore reduces significantly the computation load and the influence of background pixels. Moreover, the proposed method employs multiple color channels for finding edge pixels and their orientations

Table 4
Lane detection performance of the proposed method on the training set for different λ.

Parameter λ	1	2	3	4	5	6	7	8	9	10
Recall (%)	91.0	92.2	92.6	92.8	93.2	93.2	93.2	93.2	93.2	93.2
Precision (%)	97.0	96.2	96.3	96.3	96.3	96.3	96.1	96.1	95.9	95.9
F-measure (%)	93.9	94.2	94.4	94.5	94.7	94.7	94.6	94.6	94.5	94.5

Fig. 6. Visual results of vanishing point estimation. Ground-truth VP: *red* dot. VP detected by the proposed method: *green* marker. VP detected by Hough-based method [27]: *yellow* marker. VP detected by Gabor-based method [16]: *blue* marker. VP detected by OLDOM [30]: *cyan* marker. See the electronic color image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

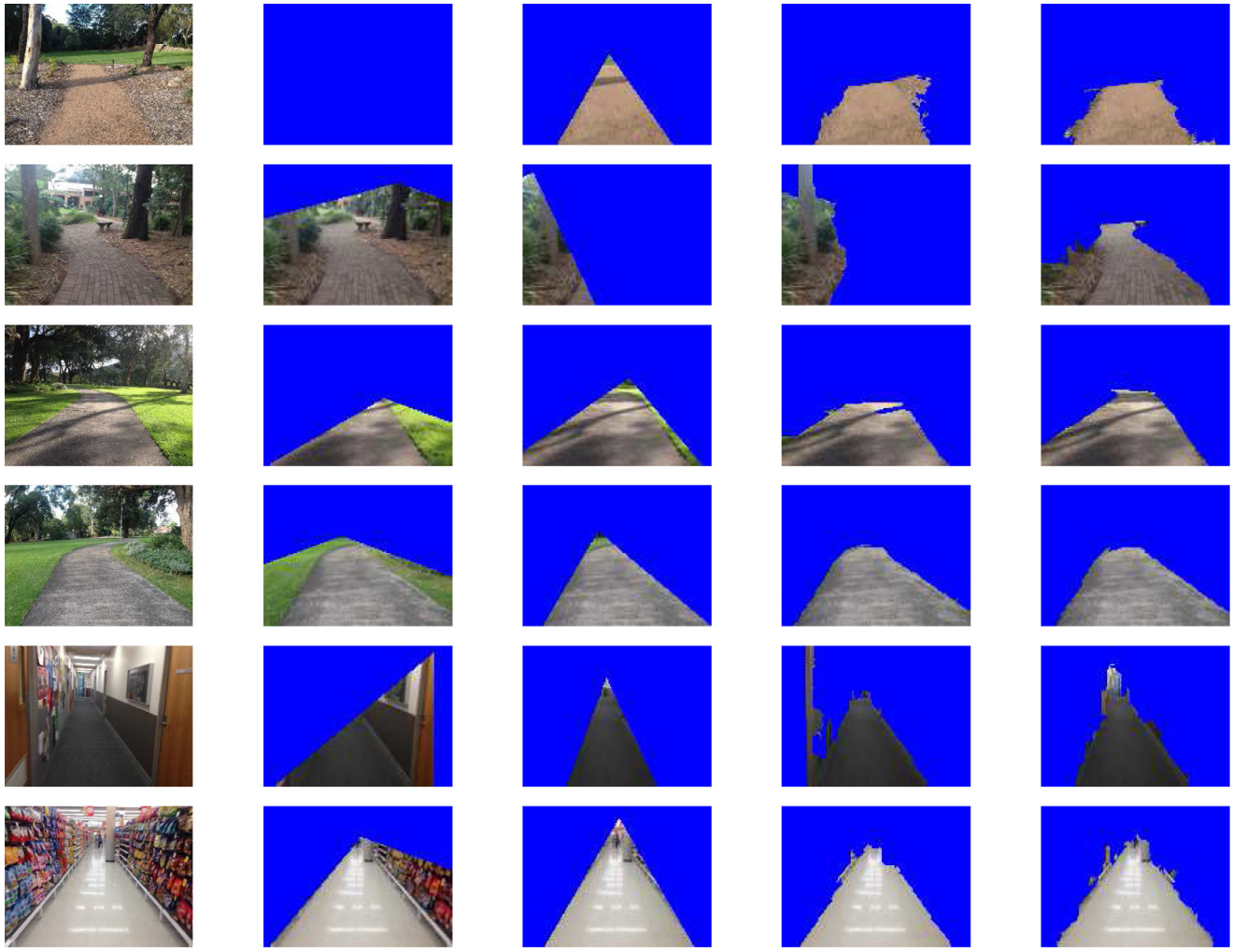

Fig. 7. Visual comparative results of different methods for pedestrian lane detection. Column 1: input images. Column 2: output of the edge-based method [16]. Column 3: output of the lane-border detection method [26]. Column 4: output of the proposed method using the RGB color space. Column 6: output of the proposed method using the IIS color space. See the electronic color image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(via color tensor). Hence, it can distinguish color pixels even if they have similar intensity.

For images of size 100 × 140 pixel, the average processing time per image of the proposed method (0.595 s) was significantly shorter than that of the Gabor-based method (2.980 s) and was almost the same as that of the OLDOM (0.595 s). That is, the proposed method was about 5.0 times faster than the Gabor-based method. Although the Hough-based method had the shortest processing time per image (0.029 s), it also had the highest error (0.1199) among the four tested methods. Fig. 6 shows examples of VPE using different methods. As can be seen, the proposed method estimates the vanishing point more accurately, compared to the Hough-based method, Gabor-based method, and the OLDOM.

4.4. Analysis of pedestrian lane detection

For pedestrian lane detection, we compared the proposed method with two related methods:

- *Edge-based method* (Kong et al. [16]): This approach detects the lane boundaries from edges directed towards the vanishing point, using the color and orientation features of lane borders. In the experiments, we used the MATLAB code provided by the authors of [16], and adjusted it using the training data to suit better this application.
- *Lane-border detection method* (Le et al. [26]): This method is our previous work, and it finds two lane borders among the edges pointing to the vanishing point. Each edge is represented by

Table 7
Performance comparison of pedestrian lane detection algorithms on the test set.

Methods	Recall (%)	Precision (%)	F-measure (%)	Processing time (s)
Edge-based method (Kong et al. [16])	62.7	65.0	63.8	3.04
Lane-border detection method (Le et al. [26])	89.0	89.7	89.3	1.20
Proposed method using RGB	92.1	94.9	93.5	1.94
Proposed method using IIS	**93.5**	**97.2**	**95.3**	**2.81**

Fig. 8. Visual sample results of the proposed method for detecting pedestrian lanes in indoor and outdoor environments. Columns 1, 3, 5 and 7: input images. Columns 2, 4, 6 and 8: detected lanes. See the electronic color images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

two features: (i) the color difference between two regions adjacent to the edge, and (ii) the orientation difference of neighboring pixels to the edge. Each region formed by a pair of edges is described by two features: (i) its color uniformity, and (ii) the direction of the bisector of the edges. A pair of edges is considered as the lane borders if the likelihood of their edge and region features is the highest among all pairs. This method does not use lane segmentation technique proposed in Section 3.3.

Table 7 shows the performance of different methods for pedestrian lane detection on the test set of 1500 images. Using the RGB color space, the proposed method had a recall rate of 92.1%, a precision rate of 94.9%, and an F-measure of 93.5%. Using the IIS color space, it achieved a recall rate of 93.5%, a precision rate of 97.2%, and an F-measure of 95.3%.

The proposed method outperformed the edge-based method [16], which had a recall rate of 62.7%, a precision rate of 65.0% and an F-measure of 63.8%. The edge-based method uses only the color and orientation properties of lane borders, and it is therefore susceptible to background edges. In contrast, the proposed method employs the properties of not only lane borders but also lane regions (appearance and shape).

The proposed method also had better recall and precision rates than the lane-border detection method [26] (recall rate of 89.0%, precision rate of 89.7% and F-measure of 89.3%). The lane-border detection method finds the lane borders from edges pointing to the vanishing point, and hence it only detects straight lanes or the straight part of curved lanes.

Fig. 7 shows pedestrian detection results of different methods. The results show the robustness and effectiveness of the proposed method compared with the previous methods [16,26]. These results also demonstrate that the proposed method using the IIS color space is more robust than using the RGB color space.

Table 7 also shows the average processing time per test image of the lane detection methods. These processing times were recorded for MATLAB implementation and an image size of 100 $\times$ 140 pixel on a PC with 3.4 GHz CPU. The proposed method (average time 2.81 s) was 1.08 times faster than the edge-based method [16] (average time 3.04 s). The proposed method was 2.60 times slower than the lane-border detection method (average time 1.20 s). Note that the proposed method required only 0.95 s on average to find the lane border (vanishing point estimation and sample region selection); it required the extra time of 1.86 s for lane segmentation (i.e. the processing steps described in Section 3.3). Nevertheless, the processing speed of the proposed method is an aspect that needs to be improved in the future.

Several outputs of the proposed method are shown in Fig. 8. There are some segmentation errors, e.g. in (Row 2, Column 2) where there is a strong shadow. However, in most cases the lane is segmented correctly. In summary, the experimental results presented in this section have shown that the proposed method can detect pedestrian lanes with various surfaces, under different imaging conditions.

5. Conclusion

This paper presents a method for pedestrian lane detection in unstructured environments, by combining color, edge, and shape features. The proposed method uses the vanishing point to automatically determine a sample lane region, from which a lane model is adaptively constructed. Evaluation results on a large data set have shown that the proposed method is able to detect various types of unstructured pedestrian lanes, in outdoor and indoor scenes under challenging environmental conditions. It also has higher accuracy compared to two other existing pedestrian lane detection methods. The paper also presents an efficient and accurate method based on the color tensor for vanishing point estimation. The proposed methods for vanishing point detection and pedestrian lane detection can have several applications, such as assistive navigation for vision-impaired people, intelligent wheelchairs, autonomous robots or vehicles operating on open roads.

References

[1] A.J. Jackson, J.S. Wolffsohn, Low Vision Manual, Elsevier, Philadelphia, USA, 2007.
[2] L.E.F. Simpson R.C., R.A. Cooper, How many people would benefit from a smart wheelchair? J. Rehabil. Res. Dev. 45 (1) (2008) 53–72.
[3] J. Kim, H. Shin, Algorithm and SoC Design for Automotive Vision Systems, Springer, 2014.
[4] I. Ulrich, J. Borenstein, The GuideCane – applying mobile robot technologies to assist the visually impaired, IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. 31 (2) (2001) 131–136.
[5] S. Shoval, J. Borenstein, Y. Koren, Auditory guidance with the Navbelt – a computerized travel aid for the blind, IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. 28 (3) (1998) 459–467.
[6] HumanWare, BrailleNote GPS, 2015, URL: store.humanware.com/hus/braillenote-gps-software-and-receiver-package.html (accessed 31.12.15).
[7] S. Se, M. Brady, Road feature detection and estimation, Mach. Vis. Appl. 14 (3) (2003) 157–165.
[8] M.S. Uddin, T. Shioyama, Bipolarity and projective invariant-based zebra-crossing detection for the visually impaired, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2005, pp. 22–30.
[9] V. Ivanchenko, J. Coughlan, S. Huiying, Detecting and locating crosswalks using a camera phone, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–8.
[10] M.C. Le, S.L. Phung, A. Bouzerdoum, Pedestrian lane detection for assistive navigation of blind people, in: Proceedings of International Conference on Pattern Recognition, 2012, pp. 2594–2597.
[11] J.D. Crisman, C.E. Thorpe, SCARF: a color vision system that tracks roads and intersections, IEEE Trans. Robot. Autom. 9 (1) (1993) 49–58.
[12] C. Tan, H. Tsai, T. Chang, M. Shneier, Color model-based real-time learning for road following, in: Proceedings of IEEE Conference on Intelligent Transportation Systems, 2006, pp. 939–944.
[13] Y. Sha, G.-y. Zhang, Y. Yang, A road detection algorithm by boosting using feature combination, in: Proceedings of IEEE Intelligent Vehicles Symposium, 2007, pp. 364–368.
[14] J.M. Alvarez, T. Gevers, A.M. Lopez, Vision-based road detection using road models, in: Proceedings of IEEE International Conference on Image Processing, 2009, pp. 2073–2076.
[15] C. Rasmussen, Texture-based vanishing point voting for road shape estimation, in: Proceedings of British Machine Vision Conference, 2004, pp. 470–477.
[16] H. Kong, J.Y. Audibert, J. Ponce, General road detection from a single image, IEEE Trans. Image Process. 19 (8) (2010) 2211–2220.
[17] M.A. Sotelo, F.J. Rodriguez, L. Magdalena, L.M. Bergasa, L. Boquete, A color vision-based lane tracking system for autonomous driving on unmarked roads, Auton. Robots 16 (1) (2004) 95–116.
[18] O. Ramstrom, H. Christensen, A method for following unmarked roads, in: Proceedings of IEEE Intelligent Vehicles Symposium, 2005, pp. 650–655.
[19] Y. He, H. Wang, B. Zhang, Color-based road detection in urban traffic scenes, IEEE Trans. Intell. Transp. Syst. 5 (4) (2004) 309–318.
[20] C. Oh, J. Son, K. Sohn, Illumination robust road detection using geometric information, in: Proceedings of International IEEE Conference on Intelligent Transportation Systems, 2012, pp. 1566–1571.
[21] O. Miksik, P. Petyovsky, L. Zalud, P. Jura, Robust detection of shady and highlighted roads for monocular camera based navigation of UGV, in: Proceedings of IEEE International Conference on Robotics and Automation, 2011, pp. 64–71.
[22] J.M. Alvarez, A.M. Lopez, Road detection based on illuminant invariance, IEEE Trans. Intell. Transp. Syst. 12 (1) (2011) 184–193.
[23] J.M. Alvarez, T. Gevers, Y. LeCun, A.M. Lopez, Road scene segmentation from a single image, in: Proceedings of European Conference on Computer Vision, 2012, pp. 376–389.
[24] J.D. Crisman, C.E. Thorpe, UNSCARF – a color vision system for the detection of unstructured roads, in: Proceedings of IEEE International Conference on Robotics and Automation, 1991, pp. 2496–2501.
[25] C.-K. Chang, C. Siagian, L. Itti, Mobile robot monocular vision navigation based on road region and boundary estimation, in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 1043–1050.
[26] M.C. Le, S.L. Phung, A. Bouzerdoum, Pedestrian lane detection in unstructured environments for assistive navigation, in: Proceedings of International Conference on Digital Image Computing: Techniques and Applications, 2014, pp. 1–8.
[27] Y. Wang, E.K. Teoh, D. Shen, Lane detection and tracking using b-snake, Image Vis. Comput. 22 (4) (2004) 269–280.
[28] J.P. Tardif, Non-iterative approach for fast and accurate vanishing point detection, in: Proceedings of International Conference on Computer Vision, 2009, pp. 1250–1257.

[29] F.A. Andalo, G. Taubin, S. Goldenstein, Vanishing point detection by segment clustering on the projective space, in: Proceedings of European Conference on Computer Vision, 2012, pp. 324–337.
[30] P. Moghadam, J.A. Starzyk, W.S. Wijesoma, Fast vanishing-point detection in unstructured environments, IEEE Trans. Image Process. 21 (1) (2012) 425–430.
[31] J. van de Weijer, T. Gevers, A.W.M. Smeulders, Robust photometric invariant features from the color tensor, IEEE Trans. Image Process. 15 (1) (2006) 118–127.
[32] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, Int. J. Comput. Vis. 59 (2) (2004) 167–181.
[33] T. Gevers, A.W. Smeulders, H. Stokman, Photometric invariant region detection, in: Proceedings of British Machine Vision Conference, 1998, pp. 659–669.
[34] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522.
[35] C. Chang, C. Siagian, L. Itti, Mobile robot vision navigation and localization using gist and saliency, in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010, pp. 4147–4154.
[36] M. Everingham, L. Van Gool, C.I. Williams, J. Winn, A. Zisserman, The Pascal visual object classes (VOC) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[37] P. Dollar, C. Wojek, B. Schiele, P. Perona, Pedestrian detection: an evaluation of the state of the art, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 743–761.