RANGE VIDEO SEGMENTATION

Michal Haindl, Pavel Žid, Radek Holub

Institute of Information Theory and Automation, Academy of Sciences CR
Pattern Recognition
182 08 Prague, Czech Republic

ABSTRACT

An unsupervised range video segmentation method based on a spatial probabilistic model for intended vehicle-based safety and warning system applications is introduced. Statistical range data discontinuities are represented by a wide-sense Markov model which guides the subsequent line-based region growing process. Single frame segmentations are mutually corrected using the continuity constraint. The resulting segmentation allows tracking moving objects and estimating their distance and velocity. The method is illustrated on synthetic range video data.

Index Terms–Range segmentation, Markov processes

1. INTRODUCTION

Autonomous advanced vehicle driving or safety systems in the near future will rely on range cameras to measure terrain geometry, road parameters or distance to obstacles. Such systems could either warn a driver of an imminent dangerous situation or to activate steering system, brakes, or air bags. Reliable and accurate multiple moving objects tracking and velocity estimation requires a high-speed high-resolution range camera completed with the corresponding range video segmentation software. Range images store, instead of brightness or colour information, the depth at which the ray associated with each pixel first intersects the object observed by a range camera. Range image provides geometric information about a scene objects independent of the position, direction, spectrum and intensity of light sources illuminating the scene, or the reflectance properties (with some limits) of the objects.

Recent progress of range-camera technology, is rapidly approaching a point when these cameras will be able to capture range video in applicable frame rate and frame resolution. Such range frames then enable to precisely estimate the geometry of the 3-D environment, including all three Cartesian coordinates of the points on an object. This can make [1] motion estimation and object tracking much easier and more reliable compared with using only video-intensity images. Range data provide precise measurements of the 3D environment geometry and contrary to intensity images all three objects points coordinates. Thus we can easily and accurately estimate objects movement in the 3D space.

Two classes [1] of motion-estimation algorithms for range images were published: one assumes rigid-motion surfaces [2], and the other is for moving deformable surfaces [3]. This class one can be further divided into feature-based algorithms [4] whose performance depend on the detection of reliable range image features and the establishment of interframe correspondence among them. The other is a direct area-based algorithm [2], which is more straightforward than the feature-based algorithm. The area-based method [2] assumes a rigid moving smooth surfaces so the local tangent planes can be constructed and only the motion of the sensor relative to the rigid environment has to be recovered. Several recursive modifications of this method was developed [1, 5, 6].

We propose in the following sections a rigid motion estimation method using direct area recognition for intended transportation safety applications. The method generalises our planar static range segmenter [7] for range videos, but accommodates also defected or discontinuous videos.

2. FRAME SEGMENTATION

Single frame segmentation is based on modification of our static range image segmentation method for scene with planar face objects [7]. This algorithm is used for single range video frames segmentation but newly constrained using the previous frames segmentation results and estimated movement what improves its performance and guarantees temporal segmentation consistency.

2.1. Discontinuity Detection

We assume data on some scan line through a range data space to be modeled using an adaptive regression model (1). This model uses high spatial correlation between neighbours of a predicted range pixel:

$$y_t = P^T Z_t + e_t \tag{1}$$

where $P^T = [a_1, \ldots, a_\beta]$ is the $1 \times \beta$ unknown parameter vector $\beta = card I_t$. We denote the $\beta \times 1$ data vector $Z_t = [y_{t-i} : \forall i \in I_t]^T$ with a multi-index $t = (m, n)$; y_t is a predicted range pixel value, e_t is the white noise with zero mean and unknown dispersion. I_t is some causal neighbour index shift set. The task consists in finding the conditional prediction density $p(y_t | Y^{(t-1)})$ given the known process history

Thanks to GAČR agency for funding.

978-1-4244-7166-9/10/$26.00 ©2010 IEEE

$$Y^{(t-1)} = \{y_{t-1}, y_{t-2}, \dots, y_1, Z_t, Z_{t-1}, \dots, Z_1\}$$

and taking its conditional mean estimation $\tilde{y}$ for the predicted data. Assuming normality of e_t, conditional independence between pixels and the normal prior probability form for the unknown model parameters we have shown ([8]):

$$\tilde{y}_t = E[y_t | Y^{(t-1)}] = \hat{P}_{t-1}^T Z_t \quad , \tag{3}$$

where the parametric vector estimate

$$\hat{P}_{t-1} = V_{zz(t-1)}^{-1} V_{zy(t-1)} \tag{4}$$

contains data accumulation matrices evaluated as follows:

$$\begin{aligned} V_{zz(t-1)} &= \alpha V_{zz(t-2)} + Z_{t-1} Z_{t-1}^T \quad , \\ V_{zy(t-1)} &= \alpha V_{zy(t-2)} + Z_{t-1} Y_{t-1}^T \quad . \end{aligned}$$

Above equations were modified using a constant exponential "forgetting factor" α to allow parameter adaptation. If the prediction error is larger than the adaptive threshold

$$|\tilde{y}_t - y_t| > \frac{2.5}{l} \sum_{i=1}^{l} |\tilde{y}_{t-i} - y_{t-i}| \tag{5}$$

then the pixel t is classified as an object edge pixel (a detected step discontinuity pixel). The adaptive threshold is proportional to the local mean prediction error estimation. We assume two mutually competing regression models (1) M_1 and M_2 with the same number of unknown parameters $(\beta_1 = \beta_2 = \beta)$ and an identical neighbour index shift sets I_t they differ only in their forgetting factors $\alpha_1 > \alpha_2$. The model $M_1, \alpha_1 \approx 1$ represents homogeneous image areas while the second model better represents new information coming from crossing some face borders because it allows quicker adaptation to this new information. The minimum-error predictors used in the presented algorithm can be completed as in (6),(7):

$$\tilde{y}_t = \begin{cases} \hat{P}_{1,t-1}^T Z_t & \text{if } p(M_1|Y^{(t-1)}) > p(M_2|Y^{(t-1)}) \\ \hat{P}_{2,t-1}^T Z_t & \text{otherwise} \end{cases} \tag{6}$$

where Z_t is a data vector identical to both models,

$$\begin{aligned} p(M_i|Y^{(t-1)}) &= k\,\Gamma\left(\frac{\gamma(t-1)-\beta+2}{2}\right) \\ & \qquad \frac{\lambda_{i,t-1}^{-\frac{\gamma(t-1)-\beta+2}{2}}}{|V_{i,zz(t-1)}|^{\frac{1}{2}}} \quad , \end{aligned} \tag{7}$$

k is a common constant and $\gamma(t) = \alpha_i^2 \gamma(t-1) + 1$, $\lambda_{t-1} = V_{yy(t-1)} - V_{zy(t-1)}^T V_{zz(t-1)}^{-1} V_{zy(t-1)}$. The determinant $|V_{zz(t)}|$, $\hat{P}_{t-1}^T$ as well as λ_t can be evaluated recursively (see [8]). For numerical realization of the predictor (6) see discussion in [9].

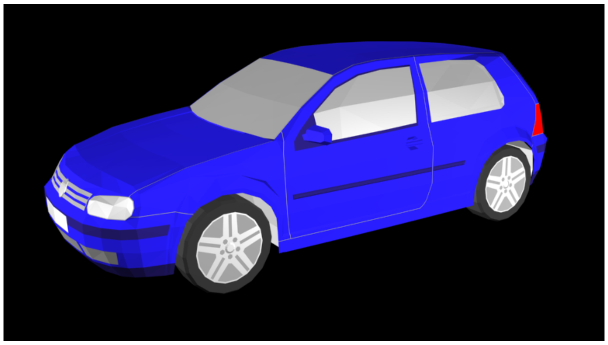

Fig. 1. 3D car model.

2.2. Planar Face Detection

The incomplete face-border information from the previous step is estimated in a line segment-based region growing process. Any line segments can serve as initial estimation but longer line segments speed up the region growing step. These lines are not allowed to cross face borders detected in the previous step of the method. A normal representing each line segment direction is computed and both normal maps (one for row-wise and one for column-wise line modeling) are low-pass filtered. Two line segments in the same column (row) are merged together iff:

1. They share one ending point.
2. The angle $(\angle())$ between their normals is less than the average angle difference between line segments already merged in this column, i.e.

$$\angle(n_l, \bar{n}_l) < \frac{k}{l} \sum_{i=1}^{l} \angle(n_{l-i}, \bar{n}_{l-i}) \quad . \tag{8}$$

The normal of the new prolonged segment is

$$\bar{n}_{l+1} = \frac{1}{t_{l+1}} \left((t_{l+1} - t_l) n_l + t_l \bar{n}_l \right) \quad , \tag{9}$$

i.e. the new prolongation influence is proportional to its length $(t_{l+1} - t_l)$. Finally we merge a line segment $\acute{l}_i$ to its neighbouring region $(\acute{l}_i \longrightarrow \acute{R}_j)$ iff:

1. An angular difference between the mean normal of lines merged into the region $\acute{R}_j$ up to now $(\acute{R}_{j,i-1})$ and the new candidate normal is less than the average normal difference during the building of this region.

$$\angle\left(n(\acute{l}_i), \bar{n}(\acute{R}_{j,i-1})\right) < \frac{k}{i-1} \sum_{l_k \in \acute{R}_{j,i-1}} \angle\left(n(\acute{l}_k), \bar{n}(\acute{R}_{j,k})\right) \tag{10}$$

2. An angular difference between the mean normal angle for a set of perpendicular line segments $(\vec{R}_i)$ crossing the candidate line $\acute{l}_i$ and the average of these normal sets for all previously merged lines to

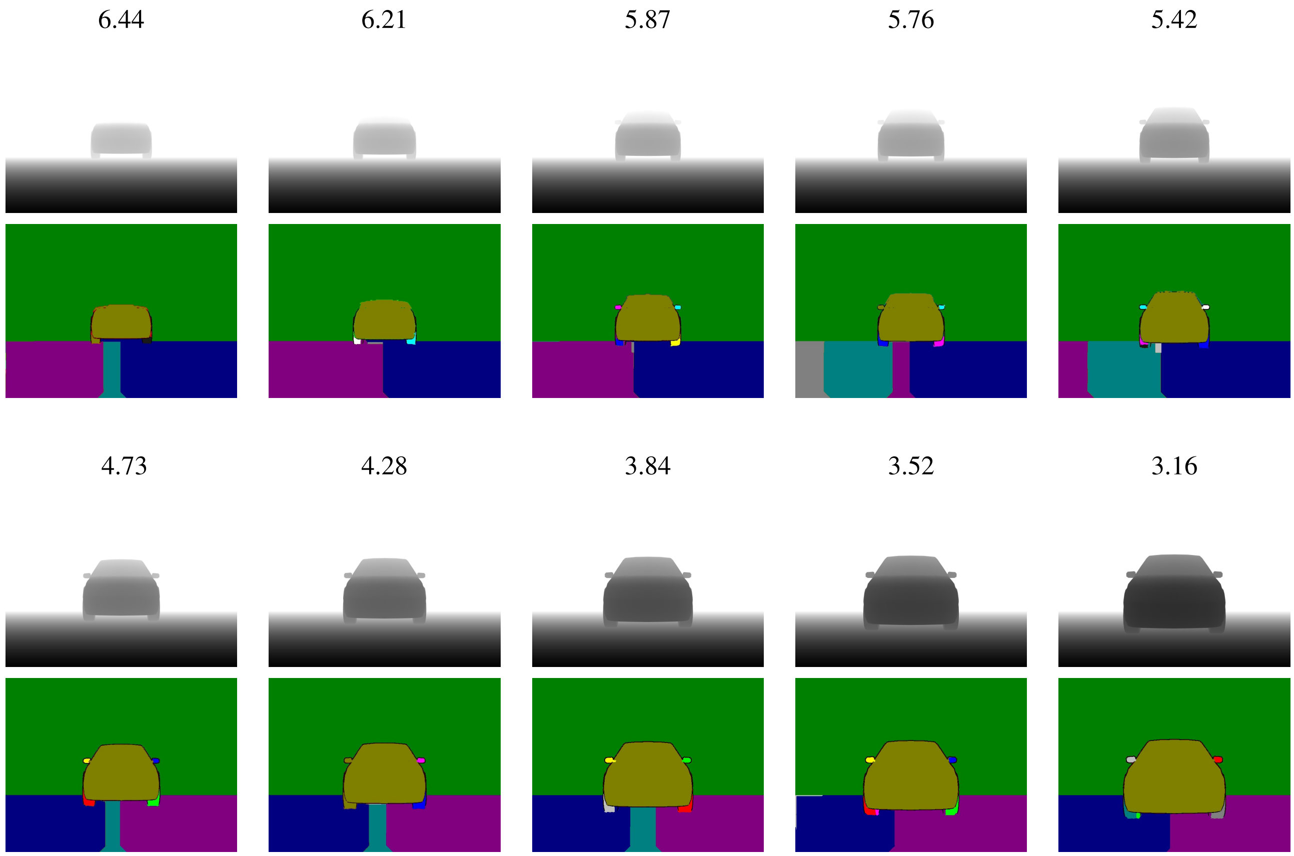

Fig. 2. Range video frames (odd image rows) their corresponding segmentation and estimated car distances from range camera in meters.

the region $\acute{R}_j$ has to be smaller than the average angular difference computed during the region $\acute{R}_j$ building process up to now.

$$\angle\left(\bar{n}(\vec{R}_i), \frac{1}{i-1}\sum_{k:\, \acute{l}_k \in \acute{R}_{j,i-1}} \bar{n}(\vec{R}_k)\right) < \frac{k}{i-1}$$

$$\sum_{k:\, \acute{l}_k \in \acute{R}_{j,i-1}} \angle\left(\bar{n}(\vec{R}_k), \frac{1}{k-1}\sum_j \bar{n}(\vec{R}_j)\right) \qquad (11)$$

The mean normal angle for the set $\vec{R}_i$ containing all perpendicular line segments crossing the line $\acute{l}_i$ is:

$$\bar{n}(\vec{R}_i) = \frac{1}{l}\sum_k n(\vec{l}_k) \;\; \forall k: \; \vec{l}_k \; \cap \; \acute{l}_i \; \neq \; \emptyset \; . \qquad (12)$$

3. EXPERIMENTAL RESULTS

In order to quantitatively evaluate our proposed 3D objects tracking and motion estimation method, we have developed a method for synthetic range video generation, because we could not obtain real traffic range video. Experimental video range sequences were simulated using the Autodesk 3D Max graphical environment. Recent laser scanners still have problems to measure outside strictly controlled environment and their resolution is either very low or they cannot reach required video frame rate.

Our synthetic range videos were produced by moving a car Fig.1 3D model on virtual model road with the predefined-motion velocity. Segmentation of synthetic range video is obviously easier than a real one, but its advantage is objective ground truth available for every frame. These image frames enable us to evaluate the accuracy of estimated-motion velocity and camera distance with reference to the actual displacement parameters.

Our synthetic range videos (Fig.2) have resolution 800× 600, 30 frames per second and simulated road lengths up to 10 meters. Estimated virtual car distances for single frames in meters are denoted in Fig.2 their precision was always better than 8 centimeters in every frame.

Although synthetic range video lacks any corruptive noise present in real range measurements, the single range frame part of the underlying segmentation algorithm was rigorously evaluated using the benchmark suggested in [10]. To get comparable results with all methods surveyed in [10], we use the ground truth data, test criteria, test set of 80 range images together with the result evaluation procedure from [10]. This performance evaluation was based on region comparison using [10] criteria on two types of real range images - the laser range finder (Perceptron) and the structured light scanner (ABW) images. Our results show good performance of our modified method [7] and significant noise robustness (performs well on noisy Perceptron range maps) in comparison with several previously published leading range segmenters [11, 12] while being much faster than most of these methods.

The processing time for the presented range video segmenter on a off-the-shelf PC (2 GHz) is 2 $[s/frame]$. The algorithm can be easily paralleled to reach real time performance on recent multiple-core processors.

4. CONCLUSION

We proposed the novel efficient range video segmentation algorithm based on the combination of range profile modeling and line-based region growing. A parallel implementation of the algorithm is straightforward, every image row and column can be processed independently by its dedicated processor. Usual handicap of segmentation methods is their lot of application dependent parameters to be experimentally estimated. Our method on the other hand requires only a contextual neighbourhood selection, which can be done using Bayesian statistics of the section three. Our core algorithm demonstrates comparable segmentation quality with the algorithms in [10] while being of an order of magnitude faster than these techniques. The proposed method is adaptive, numerically robust and still with moderate computation complexity so it can be possibly used in a real time vehicle safety system.

Acknowledgements

This research was supported by grant GAČR 102/08/0593 and partially by the MŠMT grants 1M0572 DAR, 2C06019.

5. REFERENCES

[1] H. Gharavi and S. Gao, “3-D motion estimation using range data,” *IEEE Trans. Intelligent Transportation Systems*, vol. 8, no. 1, pp. 133–143, Mar. 2007.

[2] B. K. P. Horn and J. G. Harris, “Rigid body motion from range image sequences,” *CVGIP: Image Understanding*, vol. 53, no. 1, pp. 1–13, Jan. 1991.

[3] H. Spies, B. Jahne, and J. L. Barron, “Range flow estimation,” *Computer Vision and Image Understanding*, vol. 85, no. 3, pp. 209–231, Mar. 2002.

[4] Adam Hoover, Dmitry B. Goldgof, and Kevin W. Bowyer, “Egomotion estimation of a range camera using the space envelope,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 33, no. 4, pp. 717–721, 2003.

[5] H. Gharavi and S. Gao, “3-d segmentation and motion estimation of range data for crash prevention,” in *Intelligent Vehicles Symposium, 2007 IEEE*, 2007, pp. 386 – 391.

[6] H. Gharavi and S. Gao, “Recursive motion estimation of range image,” in *ICASSP 2007*, 2007, vol. I, pp. 673 – 676.

[7] M. Haindl and P. Žid, “Fast segmentation of planar surfaces in range images,” in *Proceedings of the 14th International Conference on Pattern Recognition*, A. K. Jain, S. Venkatesh, and B. C. Lovell, Eds., Los Alamitos, August 1998, pp. 985–987, IEEE.

[8] M. Haindl and S. Šimberová, *Theory & Applications of Image Analysis*, chapter A Multispectral Image Line Reconstruction Method, pp. 306–315, World Scientific Publishing Co., Singapore, 1992.

[9] M. Haindl and P. Žid, “Fast segmentation of range images,” *Lecture Notes in Computer Science*, , no. 1310, pp. 295–302, September 1997.

[10] Adam Hoover, Gillian Jean-Baptiste, Xiaoyi Jiang, Patrick J. Flynn, Horst Bunke, Dmitry B. Goldgof, Kevin Bowyer, David W. Eggert, Andrew Fitzgibbon, and Robert B. Fisher, “An experimental comparison of range image segmentation algorithms,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 673–689, July 1996.

[11] P. J. Besl and R. C. Jain, “Segmentation through variable-order surface fitting,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 2, pp. 167–192, March 1988.

[12] Xiaoyi Jiang and Horst Bunke, “Fast segmentation of range images into planar regions by scan line grouping,” *Machine Vision and App.*, vol. 7, no. 2, pp. 115–122, 1994.