© 1994 European Association for Signal Processing, Lausanne, CH

Moving Segment Detection in Monocular Image Sequences under Egomotion

Włodzimierz KASPRZAK † and Heinrich NIEMANN †‡

†*Bavarian Research Center for Knowledge–Based Systems (FORWISS), Am Weichselgarten 7, D-91058 Erlangen, E-Mail: kasprzak@forwiss.uni-erlangen.de*

‡*University of Erlangen–Nuremberg, Institute of Computer Science (IMMD 5), Martensstr. 3, D–91058 Erlangen, E-Mail: niemann@informatik.uni-erlangen.de*

Abstract. Two moving segment detection problems in image sequences of road scenes under egomotion are discussed: the classification of image contours into moving or stationary and the 3–D object motion estimation. In both cases two solution classes are compared. The approaches of the first class (bottom–up, application–independent) are searching for image feature correspondence in subsequent images. The second class approaches perform a model–based classification and object motion estimation. Hence the first problem is solved either by image motions of several contour points or by geometric vanishing point detection. After a road detection step the contours are grouped and their depth positions are hypothesized. The two approaches applied to the problem of object motion estimation use either the 2–D length change rate of a contour group or 3–D positions of corresponding groups in subsequent images.

1. Introduction

During early *image sequence* processing one primarily wants to redetect the previously detected features in the next image in order to stabilize the image description and to obtain the *visual motion* field [1]. In case of true *camera motion* in the 3–D space (also called *egomotion*), not limited to the *lateral* camera motion, the visual motion consists of two unknown components: the first one corresponds to the object motion and the second one to the egomotion. The visual motion of stationary objects is not unique – it depends from the object position relative to the camera. For a separation of true moving segments from the stationary background the visual motion part caused by the egomotion should be estimated. On the other hand for the calculation of this unknown motion part the stationary segments should be used only.

In this paper two moving segment detection problems in monocular image sequences of road scenes under egomotion are discussed and two classes of methods for solving them are compared: *bottom–up* methods based on image motion and *model–driven* methods, based on geometry detection in 2–D or 3–D space. Related works are described for example in [2] (road scene analysis), [3] (focus of expansion detection), [4] (adaptive estimation) and [5] (3–D egomotion estimation).

2. System outline

Let us consider a system for dynamic road scene analysis with a moving observer as depicted in **Figure 1**. It consists of an application–independent module for image contour detection and image motion estimation (2–D) (left bottom scheme part), that is integrated with a 2.5–D module for contour classification and road detection (top part), and they are both interacting with two model–based modules for object initialization (3–D) and tracking (4–D) (right part).

The 2–D module has been described in [6]. Its task is to detect *closed* image contours and to estimate the image motion vectors of the contours.

After the contour classification step has been finished (as described in section 3) application–specific knowledge about the scene is used next for the detection of the *road class*, road width and the observer position relative to the middle road axis.

Because the measured image motion as well as the image locations are very sensitive to the discretization error, two stabilization schemes are applied and redundant measurements are provided. A weighted averaging of individual measurements in a short sequence of up to 5 images is performed and an *adaptive* scheme is applied (by means of a recursive linear filter) for the stabilization of the image measurements in a long image sequence.

The *contour grouping* step starts with the backprojection of the contours into the 3–D space over the road plane. Then it tries to aggregate neighbour contours from the same class together. For example the search for an "obstacle" group starts with a non–stationary "road" contour on the bottom of the image. It lookes for image neighbours that are located near the first contour, when projected back to the 3–D space over the road. The group should satisfy the geometry restrictions given by the object model (i.e. width–to–height ratio).

All initialized objects are supplied to the tracking module. The relative motion of "stationary" objects is caused by the *egomotion* of the camera vehicle. The current egomotion measurement is the result of weighted averaging of individual stationary road stripe velocities. The weights correspond to the measurement variances of every object. Both stationary and true moving objects are tracked by the use of a model–to–image match (point– or edge–based) and their states are recursively updated by means of adaptive filtering methods. The tracking module is scope of the paper [7].

In the current paper the design of the *contour classification* and the *object initialization* steps is discussed in the context of the following question: how can the 3–D object recognition task in image sequences from a low–cost camera

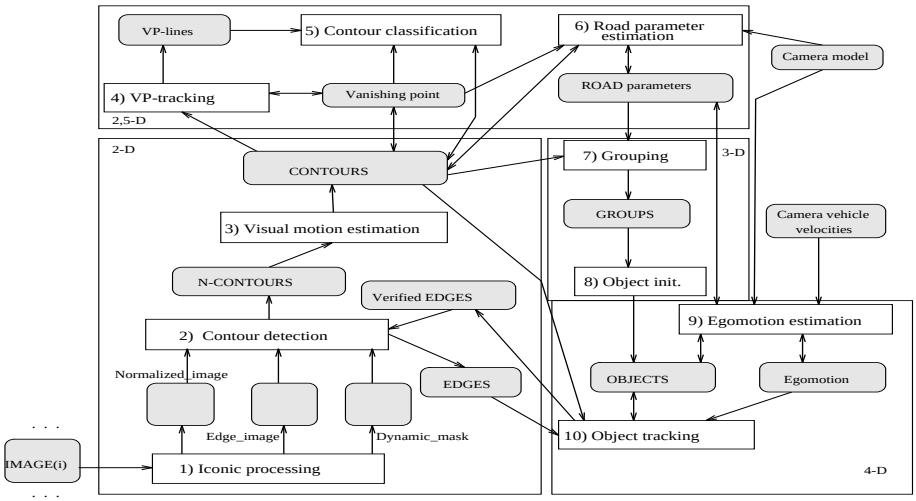

Figure 1: Processing structure of the road analysis system

be supported by the application–independent image motion of the contours. During the contour classification step the image motion in several contour points may allow the separation of moving contours from the stationary ones. In the second step the translational velocity of an object hypothesis may be initialized by using the image contour change rate instead of detecting the depth position difference of corresponding hypotheses.

3. Contour classification

3.1 VP–based contour classification

An example of the contour data results from the 2–D module is given in **Figure 2**(a). The *vanishing point* (VP) in the image is detected as the center point of an area with the highest density of hypothetic road line crossings (**Figure 2**(b). On the basis of the VP location the image pixels are classified into three classes: "road", "surrounding area" and "heaven". The contours containing some number of "road" pixels are classified as "road" contours, the contours without such pixels but containing enough "surrounding" area pixels are classified as surrounding contours. The remaining contours constitute the "heaven" (**Figure 2**(c)).

After this basic classification the "road" contours with a large amount of VP–edges (i.e. pointing towards the VP point) are classified as "road stripes" (stationary segments) and the remaining "road" contours are "obstacles" (moving segments) (**Figure 3**).

3.2 FOE–based contour classification

In the general case the existence of a vanishing point is not garanteed. In such situation an image motion–based method for moving segment detection could be used instead of the VP–based method.

In **Figure 4**(a) a closed discrete contour, its features and discrete disparity vectors are shown. The unknown motion $(\tilde{v}_x, \tilde{v}_y)$ of a continuous contour feature is approximated by a *weighted averaging* of N-1 disparity vectors for a discrete feature point – (v_x, v_y). The *weights* are directly related to the additional component of contour motion v_z – the relative contour length change rate. For the motion vector set of a true moving contour a dynamic *focus of expansion* point (C_FOE) is estimated. This is the point in an image (in general a region), where the motion lines induced by the feature motion vectors vanish.

The general method of moving segment detection is based on the distinction of individual *focus of expansion* points for each contour. The idea is to classify contours being the projections of assumed road stripes and the "surrounding" contours into *stationary* contours. The motion vectors of *stationary* contours induce then the current dynamic *focus of expansion* point (FOE). This is the center of an area in the image plane, whith the highest density of C_FOE points for stationary contours. A *moving* contour should differ from the stationary background by its visual motion (its C_FOE does not match the stationary FOE point) (**Figure 4**(b),(c)).

4. Object initialization

The initialization of an object (or generation of an object hypothesis) is equivalent to the initialization of a parametric *state vector* on the basis of one group features and model–dependent restrictions, with the knowledge about the current camera–to–road transformation. A state vctor consists of the *trajectory* and *shape* parts. The *trajectory* subvector $x(k)$ at time point t_k is a five–dimensional vector

$$x(k) = [(p_X(k), p_Z(k), \Theta(k)), (V(k), \omega(k))]^T, \qquad (1)$$

that consists of the position $(p_X(k), p_Z(k))$ relative to the camera vehicle, the orientation $\Theta(k)$ of the translational motion and the magnitudes $V(k)$ and $\omega(k)$ of translational and angular velocities.

The localization parameters (position and orientation) are model–based estimated by projecting the image group features back into the road coordinates assuming the on–road position. Then the model–based restrictions about the length-to-width and height-to-width ratios allow the orientation estimation.

The angular velocity is assumed to be equal to the current egomotion state parameter $\omega(k)$. The translational velocity along the depth axis can be hypothesized on two ways. The geometry–based method performs a short–time tracking of the object depth, whereas the application–independent

Image	$\lvert\mathcal{E}_X\rvert$	σ^2_X	$\lvert\mathcal{E}_{X*}\rvert$	σ^2_{X*}	$\lvert\mathcal{E}_Y\rvert$	σ^2_Y	$\lvert\mathcal{E}_{Y*}\rvert$	σ^2_{Y*}
	$\mathcal{X}_{VP} = x_{VP} - xo$		$\mathcal{X}^*_{VP} = x^*_{VP} - xo$		$\mathcal{Y}_{VP} = y_{VP} - yo$		$\mathcal{Y}^*_{VP} = y^*_{VP} - yo$	
1 - 5	$1.5 - 6.2$	$17 - 156$	$1.1 - 5.2$	$1 - 28$	$0.8 - 6.4$	$6 - 64$	$0.9 - 6.0$	$1 - 7$
	$\mathcal{X}_{FOE} = x_{FOE} - xo$		$\mathcal{X}^*_{FOE} = x^*_{FOE} - xo$		$\mathcal{Y}_{FOE} = y_{FOE} - yo$		$\mathcal{Y}^*_{FOE} = y^*_{FOE} - yo$	
1 - 5	$9.6 - 30.7$	$1148 - 1978$	$7.5 - 26.9$	$31 - 340$	$23.6 - 35.3$	$107 - 215$	$21.9 - 32.7$	$28 - 57$

Table 1: The average errors and the errror variances of the VP and FOE estimation in 5 image sequences

method applies the contour length change rate v_z as folllows:

$$V_Z(k) \simeq p_Z(k) \left(\frac{\gamma(k)}{1 + v_z(k)} - 1 \right) \quad (2)$$

where the coefficient $\gamma(k)$ depends from the object type and the rate $v_z(k)$.

5. Experiments

Five test sequences have been used with complexity of 100–120 contours in one image and with average contour length of 100–110 *pixels*.

5.1 FOE vs. VP–estimation

In **Table** 1 the detected errors E and the variances σ^2 of the detection (X, Y) and estimation (X^*, Y^*) of the FOE and VP points in five image sequences are given. The original VP locations (x_o, y_o) have been manually measured in the images. As the camera vehicle is moving approximately towards the VP–line, the reference value for the FOE detection error was the same as for the VP error. While comparing the errors and variances an immediate conclusion is, that the quality of VP detection is about ten times better, than the quality of FOE detection.

Thus the FOE–based approach has failed to reach his goals in practical tests. Unfortunately the differences in measured motion of different contour points have been to small for a robust and stable determination of a C_FOE point for a great number of contours.

5.2 Object motion initialization

There are two road hypotheses tracked in parallel that correspond to a 2– or 3–lane road class. The contours are grouped into 20 – 30 groups with a tracking success of 90 – 95 % . Some 10 groups correspond to a moving object hypothesis (there were 3–4 vehicles in the scene) and 10–12 groups induce a generation of up to 6 road stripes for each road hypothesis (**Figure 5**).

The **Table** 2 summarizes the results of repeated depth p_Z and object motion V initializations for a middle road stripe hypothesis in 20 images. The original values of depth p_{Zo} and velocity V_o of the moving object have been measured manually in the image sequence.

The errors of depth estimation were up to $\pm 25\%$ but the errors of translational velocity along the depth axis, calculated from the 3–D location differences, were much higher – between -57% and 67%. The quality of the same velocity, but computed by the v_z–based method, was better – errors of $\pm 17.5\%$ have been observed. There is a big contour detection instability in the image interval 7–12 as the stripe is passing a highligted area in the road. During the estimation of the v_z change rate, these errors are partly filtered out by the measurement stabilization procedures and by the redundancy of measurements (the border length and diagonal length change rates are combined).

These errors should be related to the discretization errors of object initialization in a synthesized image sequence. For a synthetic road stripe of similar size (256x256x8 bit images, contour length from 34 to 106 pixel) the detected measurement error of $v_z * l$ or of the contour center motion (v_{Cx}, v_{Cy}) was up to $0.3 pel/\delta t$ or up to 7%. At the same time the errors of individual border point motions were several times larger than this error. With known vehicle speed (ca. $1.11 m/\tau$) the error of translational velocity was below $2m$ or 7.4%.

6. Conclusion

Two classes of approaches for moving segment detection in image sequences under true camera egomotion have been compared: general image motion–based methods and model–based methods. For the first detection problem the processing results of a model–based method (VP–detection) have been of much better quality than the results of a general solution (dynamic FOE–detection). For the estimation of translational velocity along the depth camera axis a contour–length–change–based method was proposed, which is of better quality than the 3–D location difference measurement of corresponding object hypotheses.

Acknowledgements

The support from the 'Deutsche Forschungsgemeinschaft', Bonn, F.R.G., is gratefully acknowledged. The real images are by courtesy of the BMW AG., Munich, F.R.G.

References

[1] **Scott G.L. (1988)**: Local and Global Interpretation of Moving Images. *Pitman*, London.

[2] **Masaki I. (ed.) (1992)**: Vision–based Vehicle Guidance, *Springer Series in Perception Engineering*, *Springer*, New York Berlin Heidelberg.

[3] **Burger W., Bhanu B. (1988)**: Dynamic Scene Understanding for Autonomous Mobile Robots, *IEEE Conference on A.I. Applications*, *IEEE Publ.*, 736–741.

[4] **Gennery D.B. (1992)**: Visual tracking of known three-dimensional objects, *Int. Journal of Computer Vision, (7)*, 243–270.

[5] **Heeger D.I., Jepson A.D. (1992)** : Subspace Methods for Recovering Rigid Motion I: Algorithm and Implementation, *Int. Journal of Computer Vision, (7)*, 95–117.

[6] **Kasprzak W., Niemann H. (1993)**: Visual Motion Estimation from Image Contour Tracking, *Springer*, *Lecture Notes in Computer Science, (719)*, 363-370,

[7] **Kasprzak W. (1994)**: Road Object Tracking in Monocular Image Sequences Under Egomotion, *Machine Graphics & Vision, (3), No.1/2*, PAS Warsaw, 297–308.

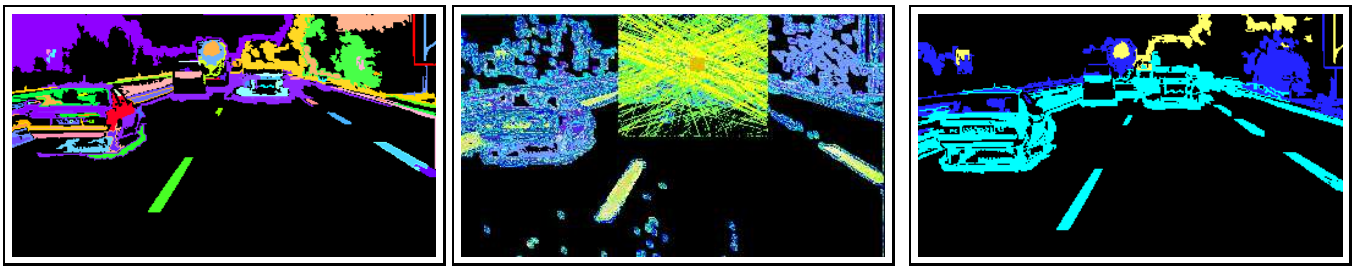

Figure 2: Basic contour classification: (a) contours, (b) vanishing point, (c) "road" and "surrounding" contours

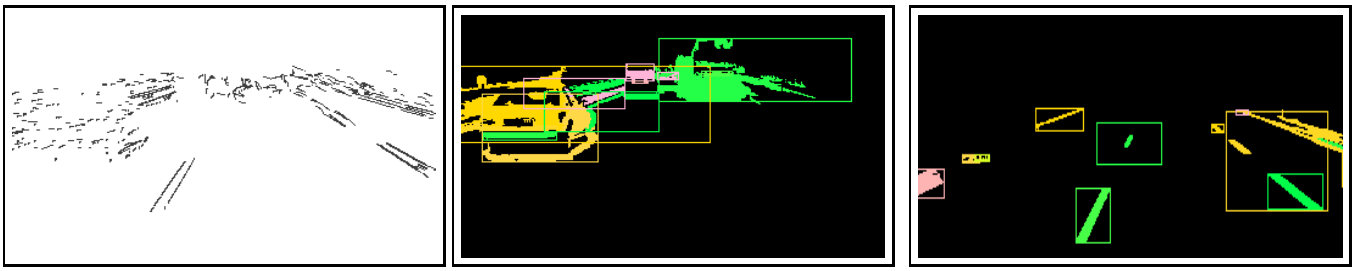

Figure 3: VP–based contour classification: (a) VP-edge detection, (b) "obstacle" contours, (c) "road stripe" contours

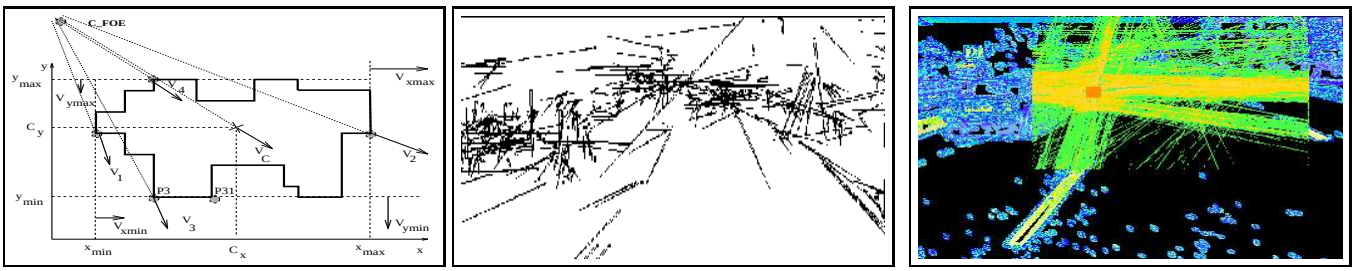

Figure 4: FOE–based contour classification: (a) the image motion of one contour and the C_FOE–point, (b) the motion vectors of the contours, (c) the detected FOE point

Space	Image feature or object state	Image $k =$						
		2	5	8	11	14	17	20
2–D	Contour length l $[pel]$	34	40	48	74	78	92	106
	Stabilized $v_z [pel/\tau]$	0.064	0.068	0.058	0.063	0.072	0.095	0.106
3–D : repeated init	VP–based depth p_Z [m]	-43.70	-39.40	-46.15	-38.95	-27.00	-25.95	-18.10
	δp_Z–based V $[m/\tau]$	1.42	1.03	0.49	1.06	1.94	1.04	1.92
	v_z–based V $[m/\tau]$	1.46	1.45	1.38	1.33	1.10	1.29	1.15
3–D : original	p_{Zo} [m]	-43.27	-39.79	-36.31	-32.84	-29.36	-25.88	-22.41
	V_o $[m/\tau]$	1.16	1.16	1.16	1.16	1.16	1.16	1.16

Table 2: Repeated object initialization (depth and translational velocity of "stationary" objects (the first middle road stripe) (p_X, Θ_S are constant and $\omega_S = 0$)) ($\tau = 0.04sec$, pel–pixel side).

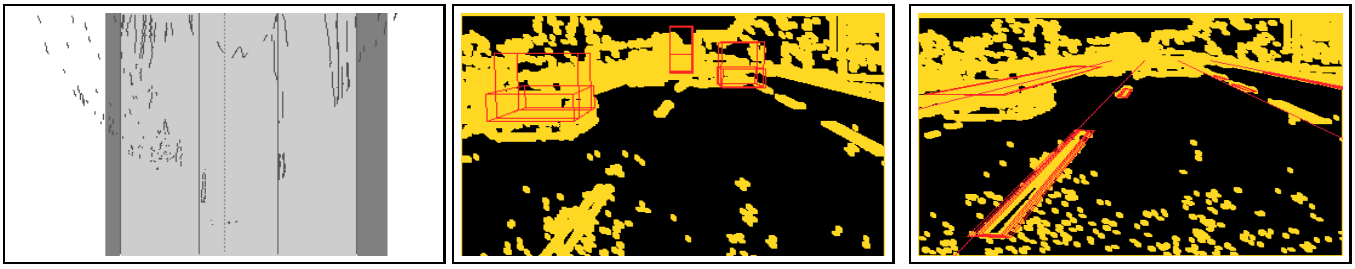

Figure 5: Object detection: (a) road lane hypothesis, (b) the car object hypotheses, (c) the road stripe objects