Motion Detection in Driving Environment Using U-V-Disparity

Jia Wang[1], Zhencheng Hu[2], Hanqing Lu[1], and Keiichi Uchimura[2]

[1] National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing 100080, China
{wangjia, luhq}@nlpr.ia.ac.cn
[2] Department of Computer Science, Kumamoto University,
Kumamoto 860-8555, Japan
{hu, uchimura}@cs.kumamoto-u.ac.jp

Abstract. Motion detection in driving environment, which aims to detect REAL moving objects from continuously changing background, is vital for *Adaptive Cruise Control* (ACC) applications. This paper presents an efficient solution for such problem using a stereovision based method. First, a comprehensive analysis about 3D global motion is given based on "U-V-disparity" concept, in which a 5-parameter model is deduced to describe global motion within U-V-disparity domain and an iterative Least Square Estimation method is proposed to estimate the parameters. Then, in order to identify separate objects, geometric analysis segments the road scene into 3D object-surfaces based on U-V-disparity features of road surfaces, roadside structures and obstacles. Finally, the motions of the segmented object-surfaces are compared with the estimated global motion to find REAL moving surfaces, which correspond to the real moving objects. The proposed algorithm has been tested on real road sequences and experimental results verified its efficiency.

1 Introduction

Motion detection in driving environment, which aims to detect REAL moving objects from continuously changing background, is vital for *Adaptive Cruise Control* (ACC) applications.

Conventionally, optical flow forms the basis of vision-based motion analysis and obstacle detection [1][2][3], which detect moving objects by measuring the flow vectors difference between the objects and background. Yet most optical flow algorithms assume that the only motion between the camera and the environment is translation, which may not be true in most driving situations.

Recent researches have paid more attention to road scene analysis based on stereovision [4][5][6][7][8][9]. Stereo analysis is the process of measuring range to an object by comparing the object projection on two or more images [9]. Based on stereo information, obstacles can be segmented by distinguishing the features belonging to them from those belonging to road surface. However, most

stereovision based method analyzes the road scene without considering the motion of cameras (*Global motion*), as well as the influence of such motion to the stereovision analysis.

In this paper, a comprehensive analysis about 3D global motion in stereovision is given based on "U-V-disparity" concept, in which a 5-parameter model is deduced to describe global motion within U-V-disparity domain and an iterative Least Square Estimation method is proposed to estimate the parameters. In order to identify separate objects, geometric analysis segments the road scene into 3D object-surfaces based on U-V-disparity features of road surfaces, roadside structures and obstacles. After that, the motions of the segmented object-surfaces are compared with the estimated global motion to find REAL moving surfaces, which correspond to the real moving objects.

2   Motion Detection Using U-V-Disparity

2.1   Stereovision

We have implemented a fast stereo module based on SSDA block matching algorithm. By careful management of cache memory and SSE technology on Pentium processor, our module can achieve real-time processing speed for dense disparity computation on a 320 by 240 image with the maximum disparity of 32 pixels. Fig. 1 shows a example of calculated disparity map.
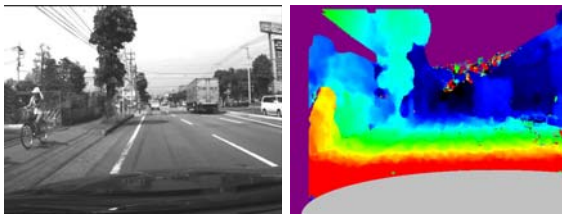
Fig. 1. Disparity map: the right figure shows the disparity map of the left image, which following a pseudo color LUT (warmer color shows a bigger value of disparity, which means closer to the observer). The grey mask area indicate "don't care" region.

2.2   Global Motion Analysis Based on U-V-Disparity

Motion detection in driving environment always involves continuously changing background caused by the motion of observer or camera itself. Such problem is called global motion in the literature.

To analyze global motion within U-V-disparity domain, we approximately assume that the stereo rig mounted on the vehicle has two coplanar cameras with the same intrinsic parameters and their horizontal co-axis is parallel to the road surface (see Fig. 2), where the pitch angle to the ground plane is θ. By putting the origin of World Coordinate System WCS (X_w, Y_w, Z_w) to the centre
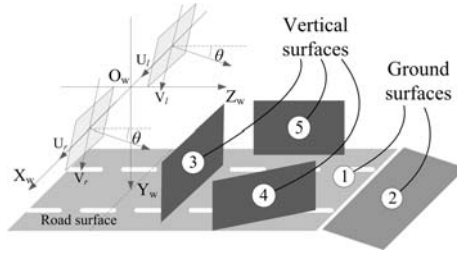
Fig. 2. Coordinate systems: The optical axis of the WCS is parallel to the ground plane and indicates the vehicle's direction of motion. The origin of camera coordinate system CCS (U, V) is put to the center of the image.

of the two stereo camera planes: O_w (as shown in Fig. 2), the transformation from WCS to CCS is achieved by:

$$\begin{cases} U_{l,r} = f \frac{X_w \pm b/2}{Y_w \sin\theta + Z_w \cos\theta} \\ V_{l,r} = f \frac{Y_w \cos\theta - Z_w \sin\theta}{Y_w \sin\theta + Z_w \cos\theta} \end{cases} \tag{1}$$

Then, disparity Δ can be deducted as:

$$\Delta = U_l - U_r = f \frac{b}{Y_w \sin\theta + Z_w \cos\theta} \tag{2}$$

Generally, global motion with respect to WCS can be described by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{3}$$

which consists of translations along X, Y, Z axis and rotations about them. In the following, global motion will analyzed within U-V-disparity domain, using the basic relationship between WCS and U-V-disparity. Note that the pitch angle θ usually remains unchanged during the vehicles/cameras are moving, which means it doesn't influence the analysis. Only concerning about the left image, We can simplify formula (1) and (2) by making $\theta = 0$ as

$$X_w = U \frac{b}{\Delta} - \frac{b}{2}, Y_w = V \frac{b}{\Delta}, Z_w = f \frac{b}{\Delta} \tag{4}$$

Based on equation (3) and (4), different global motions are analyzed separately.

1) Translation along X axis

Using WCS, translation along X axis is described by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} + \begin{bmatrix} t_1 \\ 0 \\ 0 \end{bmatrix} \tag{5}$$

Accordingly in U-V-disparity domain, such motion, combining (5) with (4), is described by

$$\begin{cases} u = u' + \Delta' \cdot t_1/b \\ v = v' \\ \Delta = \Delta' \end{cases} \tag{6}$$

2) Translation along Y axis

Similarly, translation along Y axis is described by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} + \begin{bmatrix} 0 \\ t_2 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} u = u' \\ v = v' + \Delta' \cdot t_2/b \\ \Delta = \Delta' \end{cases} \tag{7}$$

3) Translation along Z axis

Translation along Z axis is described by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ t_3 \end{bmatrix} \Rightarrow \begin{cases} u = u'/(1 + \Delta' \cdot t_3/fb) \\ v = v'/(1 + \Delta' \cdot t_3/fb) \\ \Delta = \Delta'/(1 + \Delta' \cdot t_3/fb) \end{cases} \tag{8}$$

4) Rotation about X axis

Using WCS, rotation about X axis can be described by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{9}$$

Mapping into U-V-disparity domain, there is:

$$\begin{cases} u = u'/(\cos\alpha - \sin\alpha \cdot v'/f) \\ v = (\cos\alpha \cdot v' + f\sin\alpha)/(\cos\alpha - \sin\alpha \cdot v'/f) \\ \Delta = \Delta'/(\cos\alpha - \sin\alpha \cdot v'/f) \end{cases} \tag{10}$$

If the rotation angle α is small, there is

$$\begin{cases} u \approx u' \\ v \approx v' + f\sin\alpha \\ \Delta \approx \Delta' \end{cases} \tag{11}$$

5) Rotation about Y axis

Similarly, rotation around Y axis is described by

$$\begin{cases} u = \frac{\cos\beta \cdot u' - f\sin\beta}{\cos\beta + \sin\beta \cdot (u' - \Delta'/2)/f} \overset{\beta \to 0}{\approx} u' - f\sin\beta \\ v = \frac{v'}{\cos\beta + \sin\beta \cdot (u' - \Delta'/2)/f} \overset{\beta \to 0}{\approx} v' \\ \Delta = \frac{\Delta'}{\cos\beta + \sin\beta \cdot (u' - \Delta'/2)/f} \overset{\beta \to 0}{\approx} \Delta' \end{cases} \tag{12}$$

6) Rotation about Z axis

Rotation around Z axis is described by

$$\begin{cases} u = (u' - \Delta'/2)\cos\gamma + v'\sin\gamma + \Delta'/2 \\ v = -(u' - \Delta'/2)\sin\gamma + v'\cos\gamma \\ \Delta = \Delta' \end{cases} \tag{13}$$

Based on the above analysis of separate motions, complex global motion within U-V-disparity domain can be modeled by the combination of them.

In order to find a simple expression of global motion, we only concern about the special cases when the cameras are firmly mounted on the vehicle and share the same motion with it. In this case, the most frequent and notable global motion is *translation along Z axis* (when the car is running forward/backward). Global motion will also include slight *translation along X axis* and *rotation about Y axis* when the car is wheeling. If the car is running on a slope, there will be slight *translation along Y axis* and *rotation about X axis*. *Rotation about Z axis* is the most infrequent motion and will not be considered in the estimation.

Therefore, we use such a simple combination of separate motions to model the complex global motion within U-V-disparity domain as

$$\begin{cases} u = \frac{1}{1+\Delta' \cdot T_Z}u' + \Delta' \cdot T_X + R_Y \\ v = \frac{1}{1+\Delta' \cdot T_Z}v' + \Delta' \cdot T_Y + R_X \\ \Delta = \frac{1}{1+\Delta' \cdot T_Z}\Delta' \end{cases} \tag{14}$$

There are 5 parameters in the model. T_X, T_Y and T_Z characterized the translation motions along X, Y and Z axis. R_X, R_Y correspond to the rotation about X, Y axis.

2.3   Object-Surface Segmentation

To extract moving objects, an initial segmentation of the road scene is necessary. In this paper, the U-V-disparity method in [9] is improved for the segmentation.

ROI Extraction. From Fig. 1, it can be seen that objects appear as surfaces in the 3D disparity map. In addition, moving objects generally appear as *Vertical surfaces* (defined by Fig. 2), and exist in the region between the *Ground surfaces* and sky. We name such regions as the *Regions of Interest*(ROI). In this paper, ROI is extracted by checking the Y coordinate in WCS, by means of which regions near ($¿h_2$) or far from ($¡h_1$) the ground are removed.

$$h_1 \leq Y_w \leq h_2 \Leftrightarrow \frac{h_1}{b} \leq \frac{V}{\Delta} \leq \frac{h_2}{b} \tag{15}$$

Fig. 3(c) illustrates the ROI extraction, where pixels not satisfying equation (15) are masked by the grey, and only *Vertical surfaces* within a certain range of altitude from the ground are remained for further analysis.
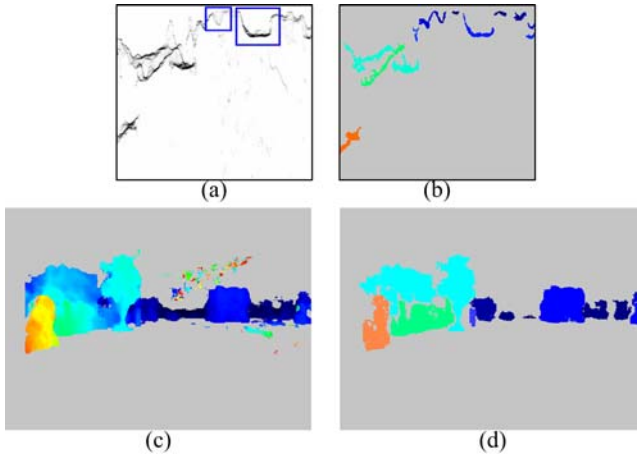
Fig. 3. ROI and object-surface segmentation: (a) U-disparity image; (b) Detected "U" curves; (c) ROI; (d) Object-surface segmentation

Object-Surface Segmentation. Following the U-V-disparity concept in [9], disparity map can be projected into U-disparity image and V-disparity image for analysis. Fig. 3(a) shows the U-disparity image of Fig. 1, in which regions with low intensity correspond to the projections of vertical surfaces. To segment the surfaces corresponding to different objects, their projections can be clustered separately by the spatial discontinuity and shape features within U-disparity image.

Since objects generally have higher disparities (closer to the observer) than the background, they always project as convex surfaces in 3D disparity maps. As the result, the convex surfaces will be projected as 2D convex curves (shapes similar to letter "U") in U-disparity iamge. An example is shown in Fig. 3(a), where the vehicles are projected as convex curves like "U" within the blue rectangles.

Based on the above analysis, we segment the surface-projections in U-disparity image as follows:

Step 1: Scanning U-Δ domain along Δ axis from Δ_{max} to 0. For each Δ, search U axis to find a *seed* pixel whose projection density is larger than a predefined threshold.

Step 2: The *seed* is expanded into a region based on the directions shown in Fig. 4. Such expanding process stops when all the candidate pixels' densities are smaller than a predefined threshold.

Step 3: Return to *Step1* to find a new *seed*.

Step 4: When the entire domain was scanned, several regions are clustered separately by the *seed*s and the expanding process. After a post-process of merging small regions to their adjacent large regions, the resultant regions in U-disparity map will be regarded as the separate projections of object-surfaces.
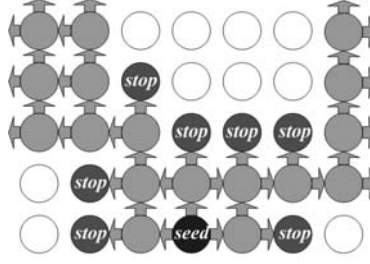
Fig. 4. Region expanding

The clustered regions are shown in Fig. 3(b). Each region is characterized by the disparity of *seed* following the pseudo color LUT. When such regions are mapped back into 3D disparity map, the segmentation result of object-surfaces is shown in Fig. 3(d).

2.4   Moving Object Detection

Iterative Least-square Estimation (ILSE) method is used to estimate the global motion parameters: First, corners are tracked through successive frames to find their correspondences. Suppose there are N corners, let (u_k, v_k, Δ_k) be the U-V-disparity coordinates for a corner k in the current frame. Its correspondence in the previous frame is (u'_k, v'_k, Δ'_k). Then, based on the N corner-pairs, ILSE method can compute the parameters as follows

$$T_Z = \frac{\sum \Delta' - \sum \Delta}{\sum \Delta \Delta'} \qquad (16)$$

$$T_X = \frac{N \sum u\Delta' + \sum Zu' \sum \Delta' - N \sum Zu'\Delta' - \sum u \sum \Delta'}{N \sum (\Delta')^2 - (\sum \Delta')^2} \qquad (17)$$

$$R_Y = \frac{(\sum u - \sum Zu') \sum (\Delta')^2 + (\sum Zu'\Delta' - \sum u\Delta') \sum \Delta'}{N \sum (\Delta')^2 - (\sum \Delta')^2} \qquad (18)$$

$$T_Y = \frac{N \sum v\Delta' + \sum Zv' \sum \Delta' - N \sum Zv'\Delta' - \sum v \sum \Delta'}{N \sum (\Delta')^2 - (\sum \Delta')^2} \qquad (19)$$

$$R_X = \frac{(\sum v - \sum Zv') \sum (\Delta')^2 + (\sum Zv'\Delta' - \sum v\Delta') \sum \Delta'}{N \sum (\Delta')^2 - (\sum \Delta')^2} \qquad (20)$$

where

$$Z = 1/(1 + \Delta' T_Z) \qquad (21)$$

Note that in the above equations, the subscript k is omitted for simplification purpose.
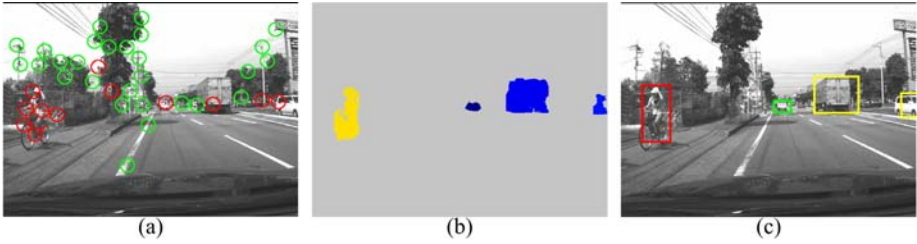
Fig. 5. Moving objects detection: (a) Detected corners and their 2D motion vectors, where corners matching the estimated global motion field are colored as green, and those who don't match are colored as red; (b) Detected moving objects; (c) The color of the rectangles indicates the distance between observer and detected objects, where *red*, *yellow* and *green* correspond to *close*, *middle* and *far*

To eliminate the influence of moving objects, the estimation procedure is performed iteratively. During each iteration, the estimated $(T_Z, T_X, T_Y, R_X, R_Y)$ are used to construct a global motion field. Such field will be compared with the U-V-disparity of each corner-correspondence, and those who do not match with the current field will be discarded. Here "match" means that U-V-disparity lies within a threshold distance from the corresponding global motion field. After that, the remaining corners are used to re-estimate $(T_Z, T_X, T_Y, R_X, R_Y)$ and enter a new iteration. Using such iterative scheme, those corners who don't follow global motion will be removed gradually, and after several iterations, the estimated parameters will converge to the final results.

Based on the segmented object-surfaces and estimated global motion parameters, moving objects (surfaces) are detected as follows: An object-surface, if
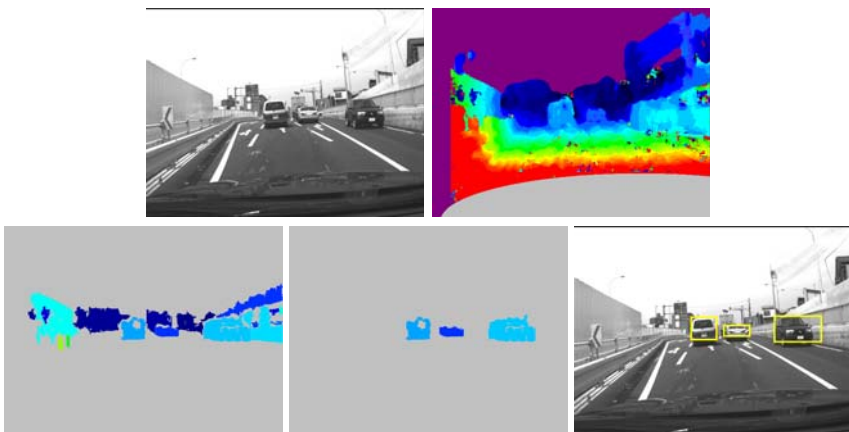

Fig. 6. Experimental results on traffic image: a road sequence involving slight rotation angles and translations about X and Y axis in global motion when the vehicle is uprising and turning left.

most of its including pixels follow the global motion in equation (14) with the estimated parameters, will be regarded as background. Otherwise, it will be regarded as having local motions and be extracted as real moving objects. Fig. 5 shows the detected moving objects of Fig. 1.

3   Experiments

The presented method has been tested on various road sequences. This section gives another two experimental results in Fig. 6 and Fig. 7. Segmentation result in Fig. 6 verifies that presented algorithm can properly handle slight rotations and translations to detect the real moving objects. In Fig. 7, the proposed algorithm makes use of equation (15) to successfully eliminate the bridge and extract the on-road objects.
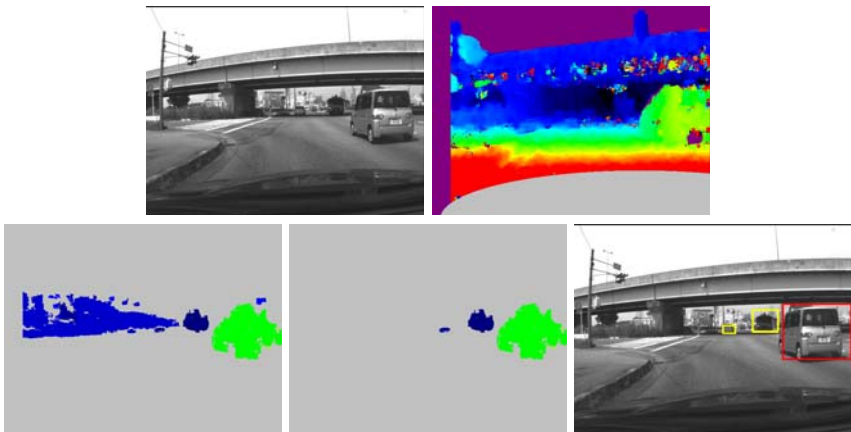

Fig. 7. Experimental results on traffic image: a road scene including an over-pass bridge, which may be easily confused as obstacles and disturb the real objects detection

4   Conclusion

Detecting REAL moving objects from moving camera is a difficult task, especially for the applications in driving environment. This paper presents an efficient algorithm for motion detection in driving environment based on stereovision analysis. On the one hand, a 5-parameter model is deduced to describe 3D global motion in stereovision using "U-V-disparity" concept, based on what an iterative Least Square Estimation method is proposed to estimate the parameters. Then, the road scene is segmented into 3D object-surfaces corresponding to separate objects based on geometric features in U-V-disparity domain. Finally, the motions of the segmented object-surfaces are compared with the estimated global motion to find REAL moving object-surfaces, which correspond to the real

moving objects. The proposed algorithm has been tested on real road sequences and experimental results verified its efficiency.

Acknowledgements

References

1. Nelson, R., Aloimonos, J.: Using flow field divergence for obstacle avoidance: towards qualitative vision. International Conference on Compute Vision (1988)
2. Enkelmann, W.: Obstacle detection by evaluation of optical flow fields form image sequence. Image and Vision Computing (1991)
3. Young, M., Hong, T., Yang, A.: Obstacle detection for a vehicle using optical flow. SAE Intelligent Vehicle (1992)
4. Luong, Q., Weber, J., Koller, D., Malik, J.: An integrated stereo-based approach to automatic vehicle guidance. International Conference on Computer Vision (1995)
5. Bertozzi, M., Broggi, A., Fascioli, A., Nichele, S.: Stereo vision based vehicle detection. IEEE Intelligent Vehicle Symposium (2000)
6. Talukder, A., Manduchi, R., Rankin, A., Matthies, L.: Fast and reliable obstacle detection and segmentation for cross-country navigation. IEEE Intelligent Vehicle Symposium (2002)
7. Labayrade, R., Aubert, D., Tarel, J.: Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation. IEEE Intelligent Vehicle Symposium (2002)
8. Sole, A., Mano, O., Stein, G., Kumon, H., Tamatsu, Y., Shashua, A.: Solid or not solid: Vision for radar target validation. IEEE Intelligent Vehicle Symposium (2004)
9. Hu, Z., Uchimura, K.: U-v-disparity: An efficient algorithm for stereovision based scene analysis. IEEE Intelligent Vehicle Symposium (2005)