Computer Vision for the Car Industry

Reinhard Klette and Zhifeng Liu

Computer Science Department, CITR, The University of Auckland
Private Bag 92019, Auckland 1142, New Zealand

This report provides a brief and informal introduction into stereo and motion analysis for driver assistance. Stereo and motion analysis play a central role in computer vision [10]. Many algorithms in this field have been proposed and carefully studied; see, for example, [2, 14] and the website *vision.middlebury.edu* for stereo and optic flow algorithms.

1 Stereo Pairs and Distance Maps

In short, a stereo pair of images allows to identify pairs of corresponding points, and those allow to calculate the distance between the projected point in the three-dimensional world and the recording cameras.

Fig. 1. A stereo pair captured by two cameras, installed on the windscreen of an ego-vehicle.

The vehicle which is used to host the stereo camera platform for capturing stereo image sequences is called *the ego-vehicle*. Figure 1 illustrates an example of a stereo video sequence, recorded and processed at Daimler AG, Germany.

After years of research on ego-motion estimation [1], the automobile industry has all the tools for producing rectified (i.e., geometrically corrected) stereo image sequences.

Calculated distance values can then be visualized in a *distance map*, where gray values represent various distance levels. Figure 2 illustrates two depth maps

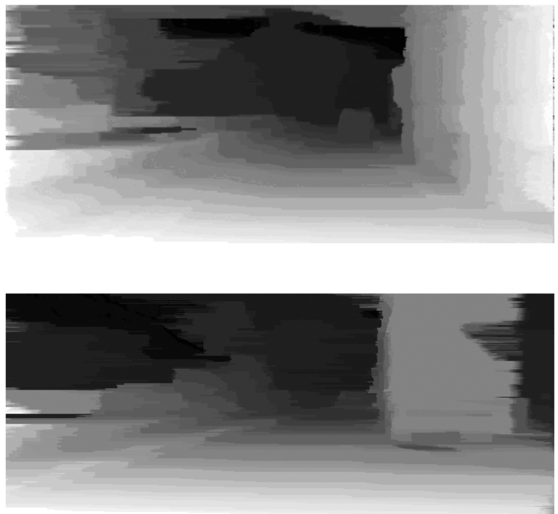

Fig. 2. Two distance maps calculated by applying a dynamic programming stereo algorithm.

as calculated in 2007 by a 4th year student (Darren Troy) at The University of Auckland. The used stereo video sequence is the same as illustrated in Figure 1.

Typically, we are interested in dense (approximate) distance maps rather than in sparse depth data. Difficulties in finding corresponding points in pairs of stereo images do have many reasons (see, for example, [10], also for some of the earlier algorithms which have been designed for stereo analysis.

We illustrate two examples of stereo analysis approaches. [4] proposed a special dynamic programming approach, basically for a pair of stereo images, in which the disparity matrix of a line is used as additional input for the calculation of the disparity map of the subsequent image line. This can now be generalized when having sequences of stereo pairs: additionally, the disparity matrix of the same line, but for the previous stereo pair, is also used. See Figure 2 for a result

Fig. 3. Edge detection followed by belief-propagation allows to calculate this dense depth map.

using this approach of propagating results within the same pair of images, and also along the time scale. This allowed for a substantial improvement.

Figure 3 illustrates another extension of a well-known approach due to the specific properties of the given image sequences. Here, belief-propagation was not performed on the given image pairs but on Sobel edge images of those. See [6] for details. See [15] for another example of a technique for calculating distance maps.

2 Downloadable Stereo Sequences

Daimler AG Germany [5] provided in 2007 seven stereo sequences for research purposes. They are captured with a calibrated pair of night-vision cameras near Stuttgart. Each sequence contains 250 or 300 frames, and features different driving environments, including highway, urban road and rural area. The ego-motion of the stereo platform has been correctly estimated and compensated [1], so that the view angle is always (about) parallel to the horizon. Furthermore, camera calibration is used for geometric rectification, such that image pairs are characterized by "standard epipolar geometry" [10]. A few "black strips" around borders of images are caused by this compensation or rectification. Figure 1 showed an example of one stereo pair in such a sequence.

The resolution of images in these sequences is 640×481. They are saved in PGM grayscale format, and available in both big endian and little endian, for convenient usage in different software, on *www.citr.auckland.ac.nz/6D/*.

3 Motion Analysis

Motion estimation for these sequences should provide information about movements of objects (speed, trajectory) as relevant for each sequence, often for identifying possible courses of conflict. Motion analysis starts in computer vision typically with optic flow calculation [**?**,8, 11], assuming that this leads to approximate calculations of the local displacement (of corresponding points between two continuous image frames). However, a few experiments with those sequences will reveal immediately the difficulty in applying optic flow algorithms successfully to those sequences, which are often blurry or fine textured (e.g., in trees).

Two typical basic assumptions behind optic flow algorithms are as follows: the brightness of the scene should be about constant, and local displacements must be small. However, both are not satisfied in those sequences. For example, the illumination changes often in one sequences, due to shading patterns of trees, or there is even different lighting for left and right camera. Object segments also move often very fast within images.

Again, sequences allow stabilization of results along the time scale (see, for example, [9]), and larger motion vectors can also be analyzed by using hierarchical approaches (see, for example, [12]). It is also recommended to apply relatively "advanced" edge detection first (such as a Canny edge detector, and not just

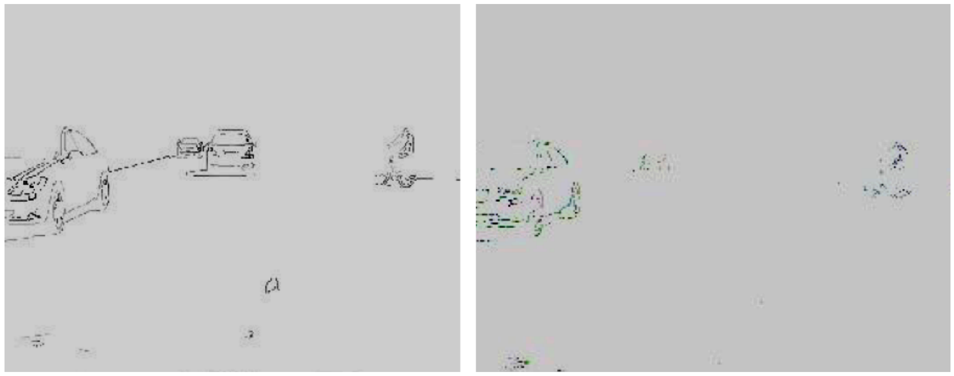

Fig. 4. Left: Canny edge image (inverted). Right: Only those edge pixels where the optic flow magnitude exceeds a threshold. Direction of optic flow is shown by hue, and length of vector by intensity (image also inverted).

the Sobel operator), and then analyze motion vectors only along those edges. See Figure 4 for an illustration (results by Xuan Guo, 4th year student in 2007 at The University of Auckland).

4 6D Analysis

Fusing stereo and motion analysis results together (i.e., 3D plus 3D, called *6D vision* in [5]), into one consistent interpretation of the scene, may allow to extract objects and their movement.

An *intersection approach* for fusion is to take all those pixels where motion information is evaluated as being reliable, and use the depth values at those pixels for combining motion and depth (or, vice-versa). However, this only allows to label very sparse pixel. Another idea is to segment depth maps, and to assign uniform motion vectors to those segments.

Fig. 5. 6D vision result, showing only moving object points which are not classified as being static background. Clustering of the shown dots allows identification of those three cars and of the bicyclist.

The use of Kalman filtering (see the book [7] or the online-tutorial [16]) is recommended to "smooth" and stabilize the movement of extracted features, and to generate more precise estimates. For example, [3] proposed this for the tracking of detected feature points, and [5] generalized this for an intersection approach.

Figure 5 illustrates a simple *background removal strategy* which uses camera calibration data which are provided for the seven test sequences (with respect to the camera's coordinate system, which is registered with respect to the car's coordinate system):

The camera's coordinate system is left handed: looking into driving direction along the z-axis, the x-axis points to the right, and the y-axis to the sky. The car coordinate system to camera coordinate system transform is the translation defined by "latpos", height, distance, followed by a rotation defined by tilt, yaw, and roll. A positive tilt means looking downward, a positive yaw means looking to the right, and a positive roll means clockwise.

Static *background* is anything what moves just (about) opposite to the movement of the host car. For calculated motion vectors, only those remain where the frame-to frame motion and the related depth information does not indicate a background situation. These are shown as dark dots in Figure 5, which resulted from a 2007 project of Xuan Guo, Zhongxia Ma, and Hao Xue, Auckland.

5 Outlook

Research on stereo sequences as described is a current hot subjects at many research centers of car companies worldwide, and also at academic institutions. See the *.enpeda..* project, *www.citr.auckland.ac.nz/projects/research/*, for example. Driver assistance based on computer vision is starting to impact safety and performance features of modern cars. Research tasks are manyfold, and will define a vivid area of research for the next couple of years.

References

1. H. Badino. A robust approach for ego-motion estimation using a mobile stereo platform. In Proc. *Int. Workshop Complex Motion*, pages 198–208, LNCS 3417, Springer, Berlin, 2006.
2. S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In Proc. *Int. Conf. Computer Vision*, to appear, 2007.
3. T. Dang, Ch. Hoffmann, and Ch. Stiller. Fusing optical flow and stereo disparity for object tracking. In Proc. *IEEE Int. Conf. Intelligent Transportation Systems*, pages 112–117, 2002.
4. S. Forstmann, Y. Kanou, J. Ohya, S. Thuering, and A. Schmitt. Real-time stereo by using dynamic programming. In Proc. *Computer Vision Pattern Recognition Workshop*, Volume 3, pages 29–36, 2004.
5. U. Franke, C. Rabe, H. Badino, and S. Gehrig. 6D-vision - fusion of stereo and motion for robust environment perception. In Proc. *DAGM*, pages 216–223, 2005.

6. S. Guan and R. Klette. Belief-propagation on edge images for stereo analysis of image sequences. In Proc. "Robot Vision 2008" (G. Sommer and R. Klette, Eds.), RobVis 2008, LNCS 4931, pages 291302, Springer-Verlag Berlin Heidelberg 2008.
7. M. H. Hayes. *Statistical Digital Signal Processing and Modeling.* John Wiley & Sons Inc., Hoboken, New Jersey, 1996.
8. B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, **17**:185–203, 1981.
9. J. Klappstein, F. Stein, and U. Franke. Detectability of moving objects using correspondences over two and three frames. In Proc. *DAGM*, pages 112–121, 2007.
10. R. Klette, K. Schlüns, and A. Koschan. *Computer Vision.* Springer, Singapore, 1998.
11. B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In Proc. *Int. Joint Conf. Artificial Intelligence*, pages 674–679, 1981.
12. N. Ohnishi, A. Imiya, L. Dorst, and R. Klette. Zooming optical flow computation. CITR-TR-197, Computer Science, The University of Auckland, Auckland, 2007.
13. J. M. Sanchiz, A. Broggi, and F. Pla. Stereo vision-based obstacle and free space detection in mobile robotics. In Proc. *Int. Conf. Industrial Engineering Appl. Artificial Intelligence Expert Systems* Volume 2, pages 280–289, 1998.
14. D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Computer Vision*, **47**:7–42, 2002.
15. T. Vaudrey, H. Badino, and S. Gehrig. Integrating disparity images by incorporating disparity rate. In Proc. "Robot Vision 2008" (G. Sommer and R. Klette, Eds.), RobVis 2008, LNCS 4931 pages 29-42, Springer-Verlag Berlin Heidelberg 2008.
16. G. Welch and G. Bishop. An Introduction to the Kalman Filter. `www.cs.unc.edu/~welch/kalman/index.html`.