Stereo Matching Based on a Combination of Simple Features Used for Matching in Temporal Image Sequences

Georg Zimmermann
Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB)
Fraunhoferstraße 1, D-7500 Karlsruhe 1, Tel. (0721)6091-255

Introduction

In image sequence understanding, crucial tasks are the extraction of information about the 3D-content and motions from the depicted scene. Further, the recognition of objects in the scene is necessary. For the first two tasks, an intermediate level of description are displacement vector fields (DVF) which describe in the stereo case the disparity between the objects depicted by the two cameras and in time sequences the displacement due to the motion of the objects in the scene and/or the camera itself. A large amount of work has been done for the estimation of DVF's; for an overview see Nagel /1/.

This paper gives a common framework for the processing of temporal image sequences and stereo pairs. Compared with most of the work done in stereo matching /2/, in this paper no a priori knowledge about the scene and no camera calibration is used. The stereo image pair used has been taken from a stereo image sequence taken under real conditions for car driving in a road scene.

Estimation of displacement vector fields

For the description of images, a multitude of features can be used, such as edges, corners, lines, blobs etc. Here, blobs delivered by the so called monotonicity operator /3/ are used. This operator detects areas of local extrema in the bandpass filtered image and represents them as the center of gravity of binary blobs. There is only one relevant parameter in this method: the center wavelength of the bandpass filter. Due to the bandpass filtering, local extrema in the image are separated at least by the distance of one wavelength, thus giving a valuable hint for the solution of the correspondence problem: If the displacement between two frames is less than half the bandpass wavelength, the phase information of the spatial Fourier components is preserved and therefore unique.

This method has been applied successfuly to a large variety of imagery of indoor and outdoor scenes. In Fig. 1, an example for an DVF is given which was used to detect automatically a starting truck in front of the camera /4/. This field displays only vectors from features, which persisted over 125 frames (5 seconds) and demonstrates besides the longevity, that vectors are found both in natural environment and on man made objects. No vectors are found in the sky region where only fuzzy results could be expected.

Correspondence in Stereo Pairs

In stereo pairs, the displacement between corresponding areas can be of an arbitrary extent. Therefore complex features of a moderate size are needed which are almost unique within the image. Such complex features are constructed here by combining blobs originating from three different wavelengths in a restricted area of the image, thus combining the phase information of three different spatial Fourier components. As these complex features have some similarity to agglomerations of more or less bright stars in the night sky, they are called 'constellations'. The construction of a constellation is done in the following way: **1)** With a large, medium and fine wavelength of the bandpass filter, extract coarse, medium and fine blobs from the image, respectively. **2)** Within a given neighborhood, assign to each coarse blob the medium ones. **3)** Do the same with the medium and fine ones. This results in a list with pointers from the coarse to the medium and from there to the fine blobs. For each blob, the center of gravity and the type of the blob (local maximum or minimum) is entered.

Two constellations are compared by searching through the lists, checking whether the medium and fine blobs are in the same position relative to the coarse one. The measure of similarity used currently simply counts the number of coinciding blobs. A correspondence is assumed if more than three blobs coincide and if the search from one constellation to the other and vice versa leads to the same combination.

Fig. 2 displays the result of a crucial test of the method. With a stereo image pair taken from a car riding with a speed of about 40 km/h through a street in the city of Karlsruhe, correspondences were looked for. No information about the relative position of the cameras was used at all. The search area for corresponding constellations was chosen to be the whole image width horizontally (512 pixels) and +/- 256 pixels vertically, which is nearly half the image area. Despite of these difficulties, 31 correspondences have been found correctly and only one is obviously false. It can be seen, that the false one can be easily eliminated by using epipolar geometry.

Conclusion

The complex image feature 'constellation' has been derived in a natural way from the simple features of the monotonicity-operator. It is demonstrated that the constellations can be used for stereo correspondence analysis. Hardware is available to compute the simple features at TV-rate /5/. Furthermore, it has been shown in /6/ that constellations are good for extracting simple image descriptions automatically.

Acknowledgement: This work was funded in parts by the German MoD.

References

/1/ H.-H. Nagel: Image sequences - ten (octal) years - from phenomenology towards a theoretical foundation. International Journal on Pattern Recognition and Artificial Intelligence; Vol. 2, No. 3 (1988), pp. 459-483.

/2/ A. Rosenfeld: Image Analysis and Computer Vision: 1988. Computer Vision, Graphics and Image Processing, 46, 169-264 (1988)

/3/ R. Kories, G. Zimmermann: A Versatile Method for the Estimation of Displacement Vector Fields from Image Sequences. Proc. of the Workshop on Motion: Representation and Analysis., May 1986, Kiawah Island Resort, Charleston, SC, pp. 101-106.

/4/ R. Kories, N. Rehfeld, G. Zimmermann: Towards Autonomous Convoy Driving: Recognizing the Starting Vehicle in Front. 9th ICPR 1988, Rome, Italy, pp. 531-535.

/5/ D. Paul, W. Haettich, W. Nill, S. Tatari and G. Winkler : VISTA: Visual Interpretation System for Technical Applications - Architecture and Use. IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol.-PAMI- 10, No. 3, May 1988, pp. 399-407.

/6/. G. Zimmermann: Creating and Verifying Automatically a Representation of Three Streets in Karlsruhe City Using Complex Image Features. Proc. Conference on Intelligent Autonomous Systems-2, Dec. 1989, Amsterdam, The Netherlands. pp. 705-714.

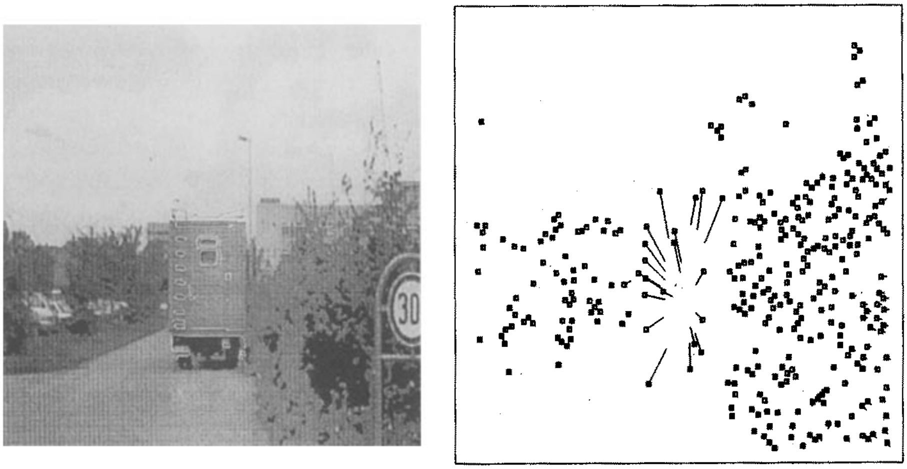
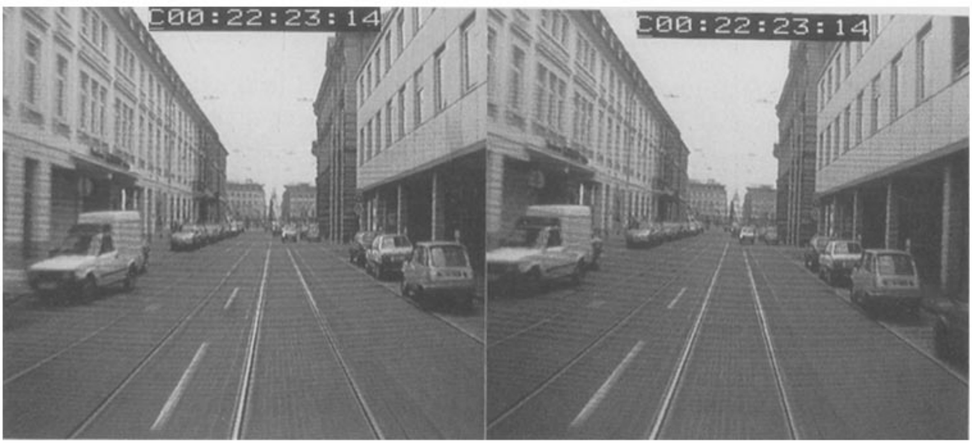


Fig. 1: Displacement vector field from a TV image sequence 5 seconds (125 frames) long. One frame is displayed on the left. On the right hand side, the displacement of only those features is indicated by needles, which could be tracked through the entire sequence. The large vectors in the center are attached to the truck which starts near the camera and becomes smaller in the image while it drives away.



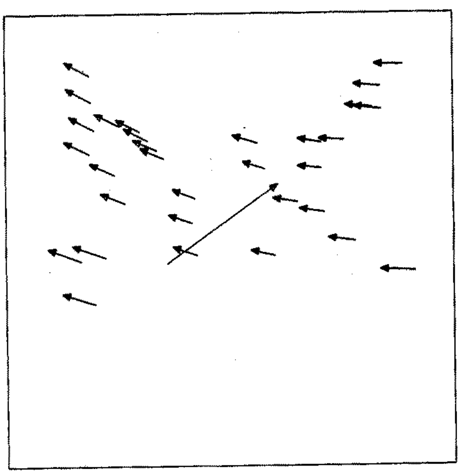


Fig. 2, Upper part: Left and right image of a stereo pair taken by cameras mounted on a moving car.
Lower part: The correspondences found automatically using the complex feature 'constellation'. It can be seen, that the two cameras are not aligned. The only false correspondence indicates the large search space (about half the image area) used in this example.