A Comparative Study on Disparity Analysis Based on Convergent and Rectified Views

O. Schreer, N. Brandenburg, P. Kauff
Image Processing Department,
Heinrich-Hertz-Institut für Nachrichtentechnik Berlin GmbH,
Einsteinufer 37, D-10587 Berlin, Germany
{schreer,brandenburg,kauff}@hhi.de

Abstract

In this paper, we present results of a comparative study on disparity analysis of convergent stereo systems. If the epipolar geometry is known, disparity analysis can be performed by computing the disparity along the epipolar line in the original views. Using a parallel stereo algorithm, the convergent views have to be rectified in order to have horizontal and parallel epipolar lines. In this study, both approaches are investigated with respect to the computational effort and the quality of the disparity analysis results.

1 Introduction

In recent years, the investigation of disparity analysis methods has mainly been concentrated on the simplified geometry of parallel stereo systems. In this case, corresponding points in two views, which refer to a common 3D point, are lying on the same horizontal scan line. Due to these simplification it was possible to develop very fast algorithms for real-time applications, like the mobile robotic or interactive multimedia services [1][2][3]. In future applications using 3D image analysis, however, parallel camera configurations will be neither practical nor possible. For example, as virtual video conferencing systems will use large displays and short viewing distances, it becomes more and more important to use highly convergent camera set-ups, where the point of convergence coincides with the location of the captured person.

One known solution to overcome this problem is to calculate parallel views virtually by applying a 2D transformation to the convergent views, the so called rectification [4][6]. These rectified views can then be analysed with a parallel matching procedure. The rectification represents a warping process where image areas are partly expanded or compressed, depending on the degree of convergence between both cameras. This geometrical distortion will influence the disparity analysis. The rectification also causes an additional computational load for each stereo image to be processed. Therefore, it can be assumed that a disparity analysis which is carried out directly at the original image grid might be advantageous over methods using rectification. In this paper, we present results of a comparative study where we have investigated these two alternative solutions on the

BMVC 2000 doi:10.5244/C.14.56

basis of computer simulations with respect to quality of the analysis results, computational complexity and real-time aspects. In the next paragraph, we briefly review the epipolar geometry and the process of rectification. A short description of the used disparity estimation algorithm will follow. Then, we explain the applications scenario which motivates us to carry out this comparative study. Finally, experimental results for both methods are presented and discussed.

2 The Epipolar Geometry and Rectification

The geometrical relation between two cameras is described by the well known fundamental matrix [7][8].

$$\mathbf{F} = \mathbf{A}_1^{-T}\mathbf{E}\mathbf{A}_2^{-1} \quad \text{with} \quad \mathbf{E} = [\mathbf{t}]_\times \mathbf{R} \tag{1}$$

Starting from a 3D point M and its projections $\mathbf{m_1}$ and $\mathbf{m_2}$ onto the two image planes I_1 and I_2, the epipolar geometry tells us that the optical ray passing through $\mathbf{m_1}$ and M is mapped onto a corresponding epipolar line $\mathbf{l_2}$ in I_2 and that therefore $\mathbf{m_2}$ must lie on $\mathbf{l_2}$ if it is visible in the second view (see **Fig. 1**). Vice versa - $\mathbf{m_1}$ necessarily lies on the complementary epipolar line $\mathbf{l_1}$ which represents the projection of the optical ray of $\mathbf{m_2}$ onto the image plane I_1**.** These relations are well described by the epipolar equation.

$$\tilde{\mathbf{m}}_1^{\,T}\mathbf{F}\tilde{\mathbf{m}}_2 = 0 \tag{2}$$

Due to this epipolar constraint, the search of corresponding points $\mathbf{m_1}$ and $\mathbf{m_2}$ can always be reduced to a 1D search along epipolar lines which are calculated as follows for one of the two available views:

$$\mathbf{l}_1 = \mathbf{F}\tilde{\mathbf{m}}_2 \quad \text{and} \quad \mathbf{l}_2 = \mathbf{F}^T\tilde{\mathbf{m}}_1 \tag{3}$$

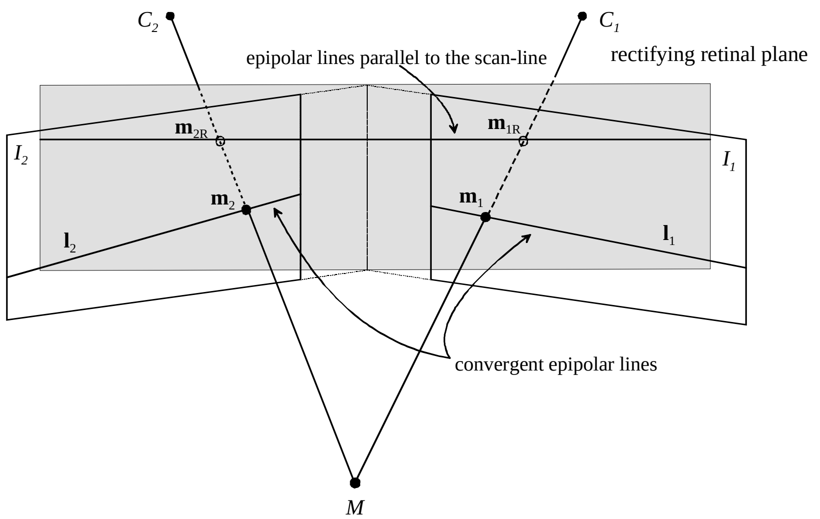

Fig. 1: The epipolar geometry and the rectified image planes

As a consequence, mature horizontal matching algorithms, which have been successfully developed for parallel stereo rigs and which are additionally able to cope with second-order problems like occlusion and disparity discontinuities, can also be exploited for dense disparity estimation in systems with arbitrary camera geometry.

In principle, this task can be approached in a twofold manner. The first approach is a 1-step solution where conventional horizontal matching methods are modified towards a direct 1-dimensional search along arbitrarily oriented epipolar lines (see section 3 for more details). The second one is a 2-step solution where both cameras are first virtually rotated until they represent a system with parallel stereo geometry. This pre-processing step is called rectification and generates rectified images with horizontal epipolar lines. Hence, point correspondences in the rectifying image planes can be searched along horizontal scan lines by using conventional matching techniques (see $\mathbf{m}_{1R}$ and $\mathbf{m}_{2R}$ in **Fig. 1**). The rectification process requires the derivation of two transformation matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ from the camera geometry. Although rectification can also be performed in the weakly calibrated case [9], we have used an approach based on the knowledge of the perspective projection matrices [11]. To obtain the transformation matrices, in this method, a number of supplementary conditions are defined, leading to a unique solution of a homogenous system of equations. The resulting matrices can then be used to transform each pixel of the original view into a point in the rectified view.

$$\tilde{\mathbf{m}}_{1r} = \mathbf{T}_1 \cdot \tilde{\mathbf{m}}_1 \text{ and } \tilde{\mathbf{m}}_{2r} = \mathbf{T}_2 \cdot \tilde{\mathbf{m}}_2 \quad (4)$$

In practice, however, it is necessary to determine grey values on the discrete grid in the rectified view. Therefore the inverse transformations $\mathbf{T}_1^{-1}$ and $\mathbf{T}_2^{-1}$ are used to map pixel positions of the rectifying image plane onto float-positions in the original images where the desired grey values can then be calculated by a bilinear interpolation.

3 The Disparity Estimation Algorithm

A hierarchical block-matching approach, providing accurate localisation of the disparities combined with high disparity resolution [3], has been used as baseline algorithm for the comparative study in this paper. The MAD (*mean of absolute differences*) has been used as a similarity measure. The algorithm is based on the following three processing steps:

- Calculation of some significant and spatially well distributed feature points on the basis of the Moravec operator [4]
- A global block-matching stage which is only applied to the few pre-selected feature points by using large search ranges as well as large measurement blocks in order to obtain stable and robust feature point correspondences (candidate vectors)
- A local block-matching stage with reduced block size measurement and limited search around neighbouring candidate vectors from the global stage in order to obtain dense disparity fields with both, highest homogeneity and best local selectivity

Finally, to quantify the reliability of the results, the consistency of left-right and right-left matches is checked on the basis of the uniqueness constraint which can be assumed in stereo (see **Fig. 2**). In the ideal case, the absolute value of the disparities of the left-right and the right-left analysis should be the same. Thus, the larger the difference Δ is, the less reliable the result. In the case of rectification, this baseline algorithm has been applied straightforwardly to the 1-dimensional search along horizontal scan lines in the rectified images. The direct search along epipolar lines, however, requires some modifications.

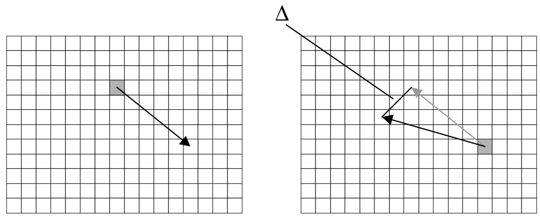

Fig. 2: Consistency check: right to left-vector (left), left to right-vector (right)

The main difference between a disparity analysis along the horizontal scan line and an arbitrary epipolar line occurs during the calculation of the position of the measurement window. In the case of arbitrary epipolar lines, this position depends on the observed point and the epipolar geometry according to eq. (3). Due to the arbitrary orientation of the epipolar line, an approximation of the line at the discrete grid is required here. To avoid floating-point operations during the calculation of the current pixel positions, the Bresenham-algorithm has been used for this purpose [1].

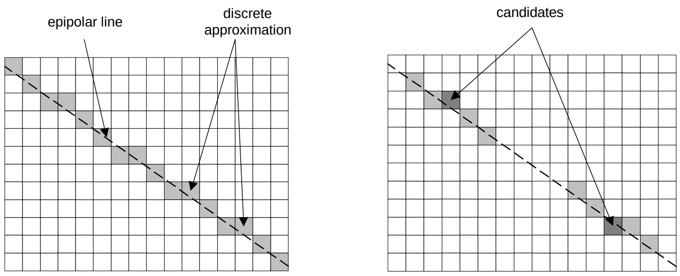

Fig. 3: Approximation of the epipolar line (left), candidates in the local stage (right)

As in the baseline algorithm, the search is done first for some distinct candidate vectors in a global stage by using a large 1-dimensional search range along the epipolar line. Then, to obtain a dense disparity field at the local stage, these candidates are checked within a limited search range with respect to the similarity and reliability.

4 The Considered Scenario

The comparative study in this paper was motivated by a very special application of the disparity analysis in the fields of future telecommunication services. The scope of this work is the implementation of a semi-immersive video conferencing system with high telepresence, providing a view-point dependent perspective for each of the conferees. One technical constraint of this telepresence system is the usage of large flat displays (e.g. plasma displays of 50 " and more). Hence, the mounting enforces a highly convergent multi-view camera set-up as, for example, shown by the sketch in **Fig. 4**. In this Starting from such a capture configuration, it becomes necessary to calculate intermediate virtual views in order to be able to offer each conferee his/her own individual perspective depending on the place where he/she is sitting at the shared virtual table.

To study the quality of novel view synthesis in such an environment, the investigations were concentrated on the left stereo pair given by the two cameras $\mathbf{C_1}$ and $\mathbf{C_2}$, because this pair represents the most convergent and - with it - the most critical set-up in the above mounting. During these investigations the position of **C_1** was fixed, whereas **C_2** was shifted horizontally along the top border of the display. While moving **C_2** from left to right, it was always looking at the same point of convergence. As a consequence, the convergence angle between the two cameras was steadily increased.

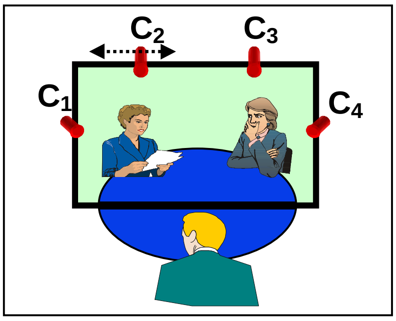

Fig. 4: Multi-view camera set-up in a virtual video conferencing system

Two sets of test images have been used for the experiments. The first was well adapted to the conference situation from **Fig. 4**. For this purpose, a puppet was placed at the default position of a conferee in front of the display and captured as described before. The original views of cameras **C_1** and **C_2** are shown in **Fig. 5**.

Fig. 5: Convergent stereo view (puppet) including epipolar lines

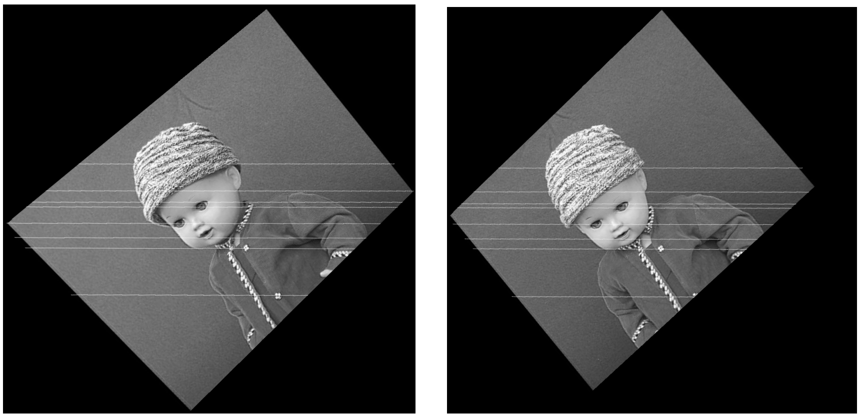

Fig. 6: Rectified views (puppet) including epipolar lines

In addition, **Fig. 6** depicts the images after the rectification process. Taking into account that only segmented video objects are processed in the above application scenario, the disparity analysis is restricted to the image area of the puppet. Furthermore, since the puppet was placed at the convergent point of the two cameras, it appears horizontally in the centre of the two images. Due to these particularities of the set-up, the area of interest (i.e. the puppet) looks quite similar in the original and the warped images. Apart from a simple rotation, possible distortions of the rectification process which may affect the quality of disparity estimation and novel view synthesis, are relatively small in this case. This is because the expected distortions mainly exist at the left and right image border rather than in the centre. Therefore, to compare the two methods under study with more generalised image material which is, for example, relevant for immersive TV applications, the second set of test images refers to a full screen shot of a toy racing circuit (see **Fig. 7**). This sequence consists of three different views, which differ in the angle of convergence around x-axis and y-axis. Two stereo pairs have been considered, which are camera 1 and 2, assigned as **position 1**, and camera 1 and 3, assigned as **position 2**. The rotation angles related to the camera coordinate system of the first camera are as follows. In camera 1, the epipolar lines of both positions are drawn.

rotation angle	φ_x	φ_y
position 1	+2°	-18°
position 2	-10°	-19°

Tab. 1. Angle of convergence around x- and y-axis of the toy racing circuit images

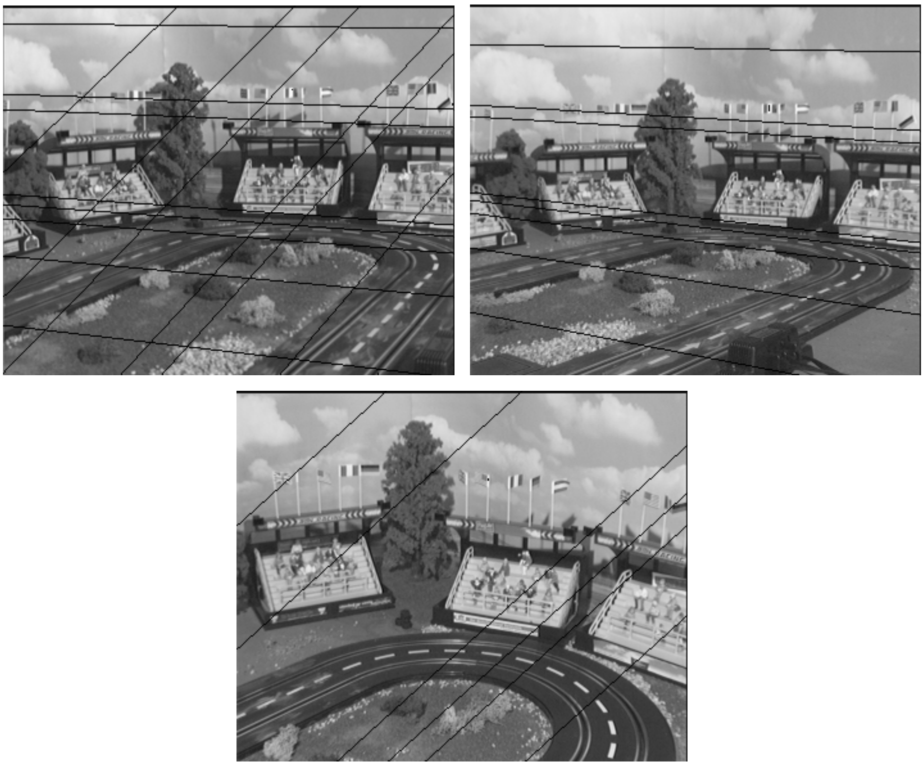

Fig. 7: Second test sequence (toy racing circuit): camera 1 - left view (top-left), camera 2 - 1st right view (top-right); camera 3 - 2nd right view (bottom), including epipolar lines

5 Experimental Results

Two different methods of disparity estimation in convergent stereo systems have been compared on the basis of computer simulations. The first one is the 1-dimensional search along arbitrarily oriented epipolar lines as described in section 3. The second one is the 2-step solution where the images are firstly warped by a rectification process (see **Fig. 6**) and then the horizontal search algorithm is applied to the scan lines of the rectified views. This comparative study was based on three different evaluation criteria: computational complexity, quality of disparity-compensated images and results of the consistency check.

3.1 Computational Complexity

Tab. 2 and **Tab. 3** show a comparison of the computational load in terms of average numbers of multiplications and additions per pixel. These figures only take into account those operations which are relevant for online processing. In particular, they do not consider those calculations which can be done off-line prior to running the real-time process due to the stationary camera configuration, such as the calculation of the epipolar line parameters in the one case or the determination of the warping coefficients of the rectification process in the other case. The total online computational costs consist of two parts: the first part is caused by the specific method, while the second part is caused by the matching process of the proposed hierarchical block-matching algorithm. The amount of computations for the matching process is similar in both methods and it is dominated by the blockwise similarity measure (*MAD*).

Both tables clearly show, that for the specific part the complexity of an estimation along epipolar lines is more than doubled compared to an estimator using rectification. The reason is that the epipolar line approach needs some additional computational load for determining the current position of the measurement block. In fact, this position has been calculated for each pixel from scratch by using the epipolar line parameter, whereas simple incrementing of the horizontal block position is sufficient for 1-dimensional block matching after rectification. But if we relate the specific amount of computations to the amount of the matching process, it can be neglected.

	Mul	Add
Rectification	8,00	20,00
Matching process	0	15750,89
Total online	**8,00**	**15770,89**

Tab. 2. Computational costs for the disparity estimation on **rectified views**

	Mul	Add
Add on by block position calculation	18,00	43,12
Matching process	0	15750,89
Total online	**18,00**	**15794,01**

Tab. 3. Computational costs for the disparity estimation on **original views**

3.1 Quality of Disparity Compensated Images

Using the results of the left-right disparity estimation, the left view can be reconstructed by the grey values of the right view. Based on this approach, a disparity-compensated image difference $I_{diff}(x,y)$ has been calculated as follows and was taken as a quality measure.

$$I_{diff}(x,y) = \left| I_L(x,y) - I_R\left(x + d_{x,L\to R}, y + d_{y,L\to R}\right)\right| \tag{5}$$

In the case of direct block matching along epipolar lines, the estimated disparities can be applied straightforwardly to eq. (5). In the case of using a rectification prior to block matching, the resulting horizontal disparities have been de-rectified and rounded to the grid of the original images. As a consequence of this post-processing of the estimated disparities, the image differences could be calculated on the basis of the original grey values. To demonstrate the synthesis results, in **Fig. 8 (left)** the synthesized left view is shown for the epipolar line estimation method. Additionally, the image difference between the original left view and the synthesized view is given in the right image.

Fig. 8: Synthesised left view (left) and image difference (right)

Fig. 9 (left) shows the results for the puppet test images where the quality measure is plotted against the convergence angle between $\mathbf{C_1}$ and $\mathbf{C_2}$ (cmp. **Fig. 4**). The drawings are very similar for both methods under study. Nevertheless, the direct epipolar line method slightly outperforms the rectification-based scheme at all convergence angles.

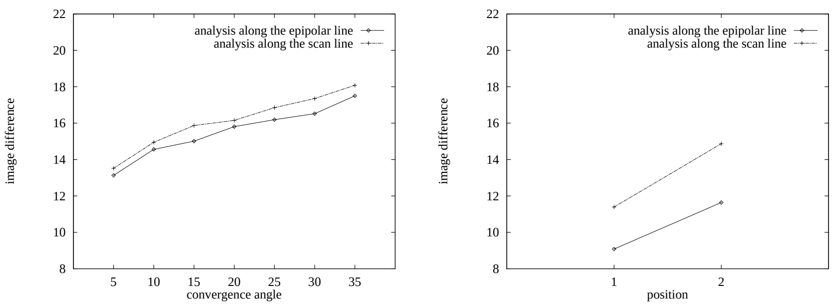

Fig. 9: Image difference for puppet sequence (left) and toy racing circuit (right)

As shown in **Fig. 9 (right)**, this benefit of the epipolar line method becomes much more significant for the second set of test images (toy racing circuit). In contrary to the puppet image, the region of interest here covers the full image frame. Thus, the fact that the distortions of rectification are stronger at the image boundary now affects the subsequent estimation process considerably. Therefore, the quality of disparity-compensated synthesised images is heavily decreased compared to the directed matching along epipolar lines.

3.1 Reliability Measure Based on Consistency Check

Similar results can be observed for the third criterion. It refers to the consistency measure Δ which has already been explained in section 3 (see **Fig. 2**). In **Fig. 10** averaged values of Δ are plotted against the convergence angle or the position. Especially, for strongly convergent set-ups the direct matching along epipolar lines is significantly more reliable than the version using rectification. The reason is decreased spatial resolution in some areas of the rectified images caused by the geometrical distortions of the warping process. Again this effect is amplified in the second set of test images (toy racing circuit) due to the higher amount of distorted areas at the image boundary (see **Fig. 10 right**).

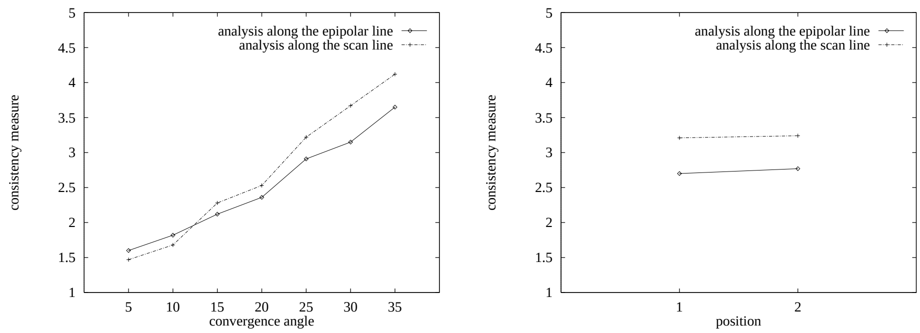

Fig. 10: Reliability measure for puppet sequence (left) and toy racing circuit (right)

6 Conclusion

In this article, two different methods of disparity estimation in highly convergent but calibrated stereo systems were compared with respect to computational load and quality of analysis results. The first method uses the rectification such that a 1-dimensional matching algorithm could be applied to the horizontal scan-lines of the rectified images. In the second method the same matching algorithm has been modified in such a way that disparities could be matched directly along arbitrarily oriented epipolar lines in the original images.

Both methods require some additional computational load for extra processes. On one hand, the rectification has to be applied. On the other hand, direct matching along epipolar lines causes some extra load for calculating the position of measurement blocks depending on the orientation of the epipolar line at the current pixel position. Theoretical investigations have shown that the extra load of the epipolar line approach is more than twice the extra load needed for rectification. However, this doubled complexity of the extra processing is marginal compared to the overall complexity of the total matching process and can therefore be neglected. The quality of the analysis results strongly depends on the type of stereo images. If segmented video objects are used, as it is the case in video conferencing applications, the results are quite similar for both methods. With respect to the mean absolute error in disparity-compensated images the epipolar line approach slightly outperforms the rectification approach for all convergence angles under inspection, but the consistency check only yields better results for large convergence angles whereas the opposite effect occurs for very small angles. Especially these latter results on consistency prove the hypothesis that the quality of analysis is affected by rectification and that this effect increases for high convergence angles. This benefit of the epipolar line method becomes much more significant in scenes where the region of interest covers the complete frame, as for example in immersive TV applications. This is because the geometrical distortion caused by rectification is much stronger at image boundaries. Thus, as a conclusion, we can summarise that the choice of the optimal method mainly depends on the following two items:

- the location of the region of interest in the image frame, because of the negative influence of rectification becomes stronger at image borders
- the degree of convergence for the same reason

For large convergence angles and full-screen video scenes, the direct epipolar line method clearly outperforms the rectification approach with respect to quality of analysis results. This gain has to be paid by a marginal and neglectable increase of complexity.

References

[1] Bertozzi M., Broggi A.: „GOLD: A Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection“, *Trans. on Image Processing*, Vol.7, No.1, January 1998.

[2] N. Grammalidis, M.G. Strintzis: „Disparity and Occlusion Estimation in Multiocular Systems and Their Coding for the Communication of Multiview Image Sequences“, *IEEE Trans. On Circuits and Systems for Video Technology*, Vol.8, No. 3, pp.328-343, June 1998.

[3] J.-R. Ohm et al: "A Realtime Hardware System for Stereoscopic Videoconferencing With Viewpoint Adaptation", *Signal Processing: Image Communication*, Special Issue on 3D Technology, Jan. 1998.

[4] H.P. Moravec: „Towards Automatic Visual Obstacle Avoidance“, *Proc. of Int. Conf. on Artificial Intelligence*, pp.584, 1977.

[5] D. Scharstein: “Stereo Vision for View Synthesis”, *IEEE Conf. On Computer Vision and Pattern Recognition*, San Francisco, pp. 852-858, June 1996.

[6] L. Falkenhagen : "Block-Based Depth Estimation from Image Triples with Unrestricted Camera Setup", *IEEE Workshop Multimedia Sig. Proc.*, Princeton, NJ, June 1997.

[7] O.D. Faugeras: “Three-Dimensional Computer Vision”, *The MIT Press*, Cambridge, Massachusetts, London, England, 1993.

[8] Z. Zhang, G. Xu: “Epipolar Geometry in Stereo, Motion and Object Recognition”, *Kluwer Academic Publisher*, Netherlands, 1996.

[9] L.Robert, C. Zeller, O.D. Faugeras, M. Hebert: „Applications of Non-Metric Vision to Some Visually-Guided Robotic Tasks“ in Y. Aloimonos (ed.), *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, pp.89-134, Lawrence Erlbaum Associates, 1997.

[10] Fusiello, E. Trucco, A. Verri: „Rectification with unconstrained stereo geometry“, *British Machine Vision Conference*, Essex, pp.400-409, Sept.1997.

[11] W.D. Fellner: Computer Grafik, *Reihe Informatik*, Band 58, BI Wissenschaftsverlag, pp.95-98, 1988.