Improving block-based disparity estimation by considering the non-uniform distribution of the estimation error

Lutz Falkenhagen and Thomas Wedi

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung,
Universität Hannover, Appelstr. 9A, 30167 Hannover, Germany
falken,wedi@tnt.uni-hannover.de

Abstract. A block-based disparity estimator is proposed that considers the non-uniform spatial distribution of the estimation error inside an image block that is mapped into another image plane using projective 2-D transformations. For this purpose, first the error variance distribution inside the image block is analytically derived. The derived error variance distribution shows four eccentric minima. As a consequence, the proposed disparity estimator arranges the image block eccentrically around the picture element (pel) to be evaluated. Thus, a reduction of the estimation error variance by a factor between 1.5 and 2 can be achieved compared to known block-based disparity estimation techniques. Since the error distribution shows four minima, four possible arrangements of the image blocks and hence four independent estimates can be made. A further reduction of the estimation error variance by a factor up to 2.6 can be achieved, when the four estimates are averaged. Additionally, an outlier detection and removal in the set of four estimates enables an increased robustness. The proposed estimator is tested using both, synthetic image pairs of known disparity and real images. The expected error reduction performance of the estimator and its increased robustness are verified.

1 Introduction

A key problem in motion and depth estimation of real objects is to find corresponding image content in a pair of images which result from a projection of the same object into both image planes [1]. Object motion leads to a *displacement* between corresponding image content in a consecutive pair of images of a monoscopic image sequence [2]. On the other hand, object depth causes a *disparity* between corresponding image content in a stereoscopic pair of images [3]. Since the estimation of disparity and displacement can be put down to the problem of finding corresponding image content, similar estimation techniques are applied.

In order to estimate displacement for a particular picture element (pel), block-based estimation approaches arrange a reference block of $N \cdot N$ pel around the pel to be evaluated. This reference block is mapped into the other image plane using a 2-D transformation and its luminance signal is compared to the luminance signal in the other image. The 2-D transformation parameters specify the displacement of the evaluated pel. They are estimated by minimising the differences between the luminance signals.

Essentially, three different 2-D transformations for displacement estimation are known from literature which differ in their number of transformation parameters:

- *translational transformation* using 2 transformation parameter [4] [5] [6]
- *affine transformation* using 6 transformation parameters [7] [8] [9] [10] [11]
- *bilinear transformation* using 8 transformation parameters [12] [13]

In general, the relation between two projections of the same physical object into two different image planes can not be modelled perfectly by a 2-D transformation of an image block. This leads to so-called model failures and thus to inaccurate estimation results. In order to determine the accuracy of displacement estimates, Buschmann recently derived a method for determining the *displacement estimation error variance* [14]. Since the real displacement vector field is unknown, the real displacement vector field is modeled as a zero-mean stochastic AR(1) process with given power spectral density. By relating the estimation results of the transformation parameters to the spatial displacement distribution inside the reference image block, an estimation error measure is derived that depends on the parameters of the stochastic process.

In [13], Buschmann extended the analysis of the displacement estimation error variance. He determined the displacement estimation error not only for the evaluated pel but also for all pels inside the reference image block depending on their position inside the block. With this extension, the *spatial distribution* of the displacement estimation error variance inside the reference image blocks is obtained. For an numerical evaluation of the spatial distribution, Buschmann empirically determined the parameters of the stochastic process by evaluating the displacement vector field of a set of videophone sequences. It turned out that the minimum of the displacement error is only in the center of the block in case of translational transformation. In case of affine and bilinear transformation, four eccentric minima were found. So far, the eccentric position of the minima is not analytically determined and not experimentally verified. All known displacement estimation approaches arrange the reference image block concentrically around the evaluated pel, where the estimation error variance is not minimal. Thus, the knowledge about the eccentric position of the minima is not yet used for an improvement of the displacement estimator.

In this paper, Buschmann's approach for calculating the displacement estimation error variance [13] is applied to disparity estimation and extended in order to improve the estimator. In case of disparity estimation, the transformation is constrained by the epipolar geometry of the stereo camera setup. Thus, the number of required transformation parameters is reduced by a factor of two. Furthermore, the stochastic properties of disparity fields are different from those of displacement vector fields. This difference is caused by the specific setup of a stereoscopic camera that introduces a disparity offset and scales the disparity variance depending on the distance between the cameras. Thus, the stochastic process has to be adapted.

In order to improve the block-based disparity estimator, the distribution of the estimation error variance is analytically derived depending on the parameters of the stochastic process. In particular, the positions of minimal estimation error variance inside the reference image block are determined. Depending on their positions, the reference image blocks are arranged eccentrically around the evaluated pel, such that an estimator with minimum estimation error variance is obtained. The decrease of the estimation error variance of the new estimator is analytically determined and verified using synthetic noisy images and real images.

In Section 2, the applied transformations are introduced. The estimation error variance distribution is specified in Section 3 for bilinear, affine and translational transformations. The proposed estimator is introduced in Section 4. Section 5 presents experimental results, followed by conclusions in Section 6.

2 Block-based disparity estimation

The task of binocular disparity estimation in a 3-D reconstruction approch is to find corresponding image points in a stereoscopic pair of images and compute the according 3-D point position by a triangulation of the lines of sight of both image points. For correspondence analysing block-based disparity estimation approaches arrange a reference block around the pel to be evaluated. This reference block is mapped into the other image plane using a projective 2-D transformation. The transformation specifies the disparity between the evaluated pel and the corresponding pel in the other image plane.

In this section, at first the applied block-based 2-D transformations with their transformation parameters are defined. Then, the estimation of transformation parameters with an Maximum-Likelihood estimator is introduced.

2.1 Definition of image block transformation

In Figure 1, a reference block and the homologous block are shown. The vector $\vec{p}_i = (x_i, y_i)^T$ denotes the position of an image point in the local coordinate system of the reference block. The vector $\vec{p}_i^h = (x_i^h, y_i^h)^T$ denotes the position of the transformed, homologous point.

A review in recent publications shows that currently mainly three different 2-D transformations are used to approximate the real mapping of image content from one image into another: translational, affine and bilinear transformation. Assuming standard stereo geometry with horizontal epipolar lines, the y-coordinate is not effected by the transformation, i.e. $y_i^h = y_i$. The mapping function of the x-coordinate is expressed by $x_i^h = \vec{B}^T \cdot \vec{A}$. In this function the vector $\vec{A}$ contains the transformation parameters and $\vec{B}$ describes the relation between the local coordinates inside the block and the transformation parameters. For the three applied 2-D transformations, the mapping functions and the vectors $\vec{A}$ and $\vec{B}$ are given in Tab. 1

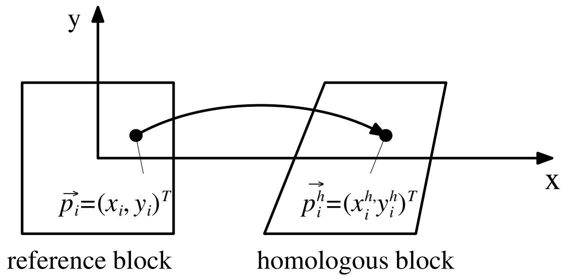


Fig. 1. Transformation of a reference block into a homologous block in case of horizontal epipolar lines.

Table 1. Definition of 2-D transformations

transformation	mapping function	$\vec{B}^T$	$\vec{A}$
translational	$x_i^h = a_1 + x_i$	$(1, x_i)$	$(a_1, 1)^T$
affine	$x_i^h = a_1 + a_2 x_i + a_3 y_i$	$(1, x_i, y_i)$	$(a_1, a_2, a_3)^T$
bilinear	$x_i^h = a_1 + a_2 x_i + a_3 y_i + a_4 x_i y_i$	$(1, x_i, y_i, x_i y_i)$	$(a_1, a_2, a_3, a_4)^T$

In the following only bilinear transformation is considered, since affine and translational transformation are subsets of it and can be easily extracted by omitting the higher order transformation parameters.

The disparity $d(\vec{p}_i)$ at position $\vec{p}_i$ inside the reference block is given by the difference between the x-coordinates of $\vec{p}_i$ and $\vec{p}_i^h$. With the parameter vector of the identical transformation

$$\vec{E} = (\,0, 1, 0, 0\,)^T, \tag{1}$$

which maps a point onto itself, the disparity can be written as a function of the used transformation and their transformation parameters

$$d(\vec{p}_i) = x_i - x_i^h = \vec{B}^T \cdot \vec{E} - \vec{B}^T \cdot \vec{A} = -\vec{B}^T \cdot \vec{\Delta a} \tag{2}$$

The differential transformation parameter vector $\vec{\Delta a}$ is given by

$$\vec{\Delta a} = \vec{A} - \vec{E} \tag{3}$$

2.2 Maximum-Likelihood (ML) disparity estimator

In order to estimate the disparity $d(\vec{p}_i)$, the differential transformation parameters $\vec{\Delta a}$ have to be estimated (2). This is done by evaluating the luminance difference between the left and right image-block [13], [15], [16]

$$fd(\vec{p}_i) = s_l(\vec{p}_i) - s_r(\vec{p}_i) = s_r(\vec{p}_i^h) - s_r(\vec{p}_i). \tag{4}$$

With the mapping functions of table 1 and a linearisation concerning the transformation parameters, equation (4) can be calculated to

$$fd(\vec{p}_i) = \sum_{k=1}^{m} \left(\frac{\partial s_r(\vec{p}_i)}{\partial a_k} \right)_{\vec{E}} \Delta a_k = \vec{H}^T(\vec{p}_i)\vec{\Delta a} \tag{5}$$

In (5) the index k denotes one of the m transformation parameters. Assuming that the observation of the luminance signal is only disturbed by Gaussian noise, the ML-estimation of the differential transformation parameters leads to the method of leads squares [15], [16]. Therefore, the estimation result is given by

$$\hat{\vec{\Delta a}} = \left(\underline{H}^T \underline{H}\right)^{-1} \underline{H}^T \cdot \vec{fd} \tag{6}$$

In $\underline{H}$ and $\vec{fd}$ the N_B observations inside the transformation block are collected:

$$\begin{aligned} \underline{H} &= \left(\vec{H}^T(\vec{p}_1), \vec{H}^T(\vec{p}_2), \cdots, \vec{H}^T(\vec{p}_{N_B})\right)^T \\ \vec{fd} &= \left(fd(\vec{p}_1),\ fd(\vec{p}_2),\ \cdots,\ fd(\vec{p}_{N_B})\right) \end{aligned} \tag{7}$$

With (6) and (2) the estimated disparity is given by

$$\hat{d}(\vec{p}_i) = -\vec{B}^T \left(\underline{H}^T \underline{H}\right)^{-1} \underline{H}^T \cdot \vec{fd}. \tag{8}$$

In order to compensate for the linearisation errors, the estimation is carried out iteratively. The estimated differential transformation parameters are accumulated over all iterations. In case of large disparities, where a differential estimation technique fails because of the assumption of a monotonic image signal, a robust disparity search technique [17] is applied for an initialisation of the disparities.

3 Local disparity estimation error variance distribution

For a derivation of the spatial distribution of the disparity estimation error inside the blocks, the disparity estimation error is defined by

$$de(\vec{p}_i) = d(\vec{p}_i) - \hat{d}(\vec{p}_i). \tag{9}$$

The stochastic properties of the disparity estimation error are described by the error variance

$$\sigma_{de}^2(\vec{p}_i) = E\left[(de(\vec{p}_i) - m_{de}(\vec{p}_i))^2\right]. \tag{10}$$

In order to compute the error variance of the estimator without knowing the real displacement vector field, Buschmann described the real displacement vector field using a stationary stochastic process. In particular an AR(1)-process is used which is described by its variance σ_{de}^2 and its autocorrelation coefficient ρ_{dd}. In contrast to displacement estimation, where a zero mean vector field is expected, an additional mean value m_d is considered in case of disparity fields since the shift between the two cameras introduces an offset. Assuming an isotropic ACF-model, the ACF of the process is defined by

$$R_{dd}(\Delta x, \Delta y) = \sigma_d^2 \rho_{dd}^{\sqrt{\Delta x^2+\Delta y^2}} + m_d^2 \tag{11}$$

With this model, Buschmann analytically derived the spatial error distribution for translational, affine and bilinear transformation. By restricting the image block transformations to horizontal epipolar geometry, the horizontal component of the displacements is turning into disparity and the vertical displacement is equal to zero. Since Buschmann assumed statistically independent displacement components, also the estimation error variance distribution of the horizontal displacement component corresponds to the one of the disparity estimate. For translational transformation the error variance is

$$\begin{aligned} \sigma_{de,d}^2(x_i, y_i) = \sigma_d^2 + m_d^2 - \frac{2}{MN} \int_{-\frac{M}{2}}^{\frac{M}{2}} \int_{-\frac{N}{2}}^{\frac{N}{2}} R_{dd}(x_i - \tau_x, y_i - \tau_x) d\tau_y d\tau_x \\ + \frac{1}{M^2N^2} \int_{-M}^{M} \int_{-N}^{N} (M - |\tau_x|)(N - |\tau_y|) R_{dd}(\tau_x, \tau_y) d\tau_y d\tau_x \end{aligned} \tag{12}$$

In case the affine transformation the error variance is

$$\begin{aligned} \sigma_{de,d}^2(x_i, y_i) = {} & \sigma_d^2 + m_d^2 \\ & - \frac{2}{MN} \int_{-\frac{M}{2}}^{\frac{M}{2}} \int_{-\frac{N}{2}}^{\frac{N}{2}} \left(1 + \frac{x_i \tau_x}{M^2/12} + \frac{y_i \tau_y}{N^2/12}\right) R_{dd}(x_i - \tau_x, y_i - \tau_x) d\tau_y d\tau_x \\ & + \frac{1}{M^2N^2} \int_{-M}^{M} \int_{-N}^{N} \Bigg((M - |\tau_x|)(N - |\tau_y|) \\ & \qquad + \frac{x_i^2[(M - |\tau_x|)^3 - 3(M - |\tau_x|)\tau_x^2]}{M^4/12} (N - |\tau_y|) \\ & \qquad + \frac{y_i^2[(N - |\tau_y|)^3 - 3(N - |\tau_y|)\tau_y^2]}{N^4/12} (M - |\tau_x|) \Bigg) \cdot \\ & \qquad R_{dd}(\tau_x, \tau_y) d\tau_y d\tau_x \end{aligned} \tag{13}$$

and for bilinear tranformation it is defined by

$$
\begin{aligned}
\sigma^2_{de,d}(x_i,y_i) = {} & \sigma_d^2+m_d^2 \\
& - \frac{2}{MN}\int\limits_{-\frac{M}{2}}^{\frac{M}{2}}\int\limits_{-\frac{N}{2}}^{\frac{N}{2}}\left(1+\frac{x_i\tau_x}{M^2/12}\right)\left(1+\frac{y_i\tau_y}{N^2/12}\right)R_{dd}(x_i-\tau_x,y_i-\tau_x)d\tau_y d\tau_x \\
& + \frac{1}{M^2N^2}\int\limits_{-M}^{M}\int\limits_{-N}^{N}\left((M-|\tau_x|)+\frac{x_i^2[(M-|\tau_x|)^3-3(M-|\tau_x|)\tau_x^2]}{M^4/12}\right)\cdot \\
& \qquad \left((N-|\tau_y|)+\frac{y_i^2[(N-|\tau_y|)^3-3(N-|\tau_y|)\tau_y^2]}{N^4/12}\right)\cdot \\
& \qquad R_{dd}(\tau_x,\tau_y)d\tau_y d\tau_x
\end{aligned}
\tag{14}
$$

These integrals give the error variance distribution inside the reference image blocks in relation to the ACF of the AR(1)-process. An evaluation of equation (14) with (11) shows, that the variance σ_d^2 of the real disparity scales the error variance σ_{de}^2 linearly. Therefore, the normalized error variance σ_{de}^2/σ_d^2 is shown in Fig. 2 in dependence on the local coordinate $\vec{p}_i = (x_i, y_i)^T$ inside the image block. A typical autocorrelation coefficient of $\rho_{dd} = 0.95$ is chosen. In addition to

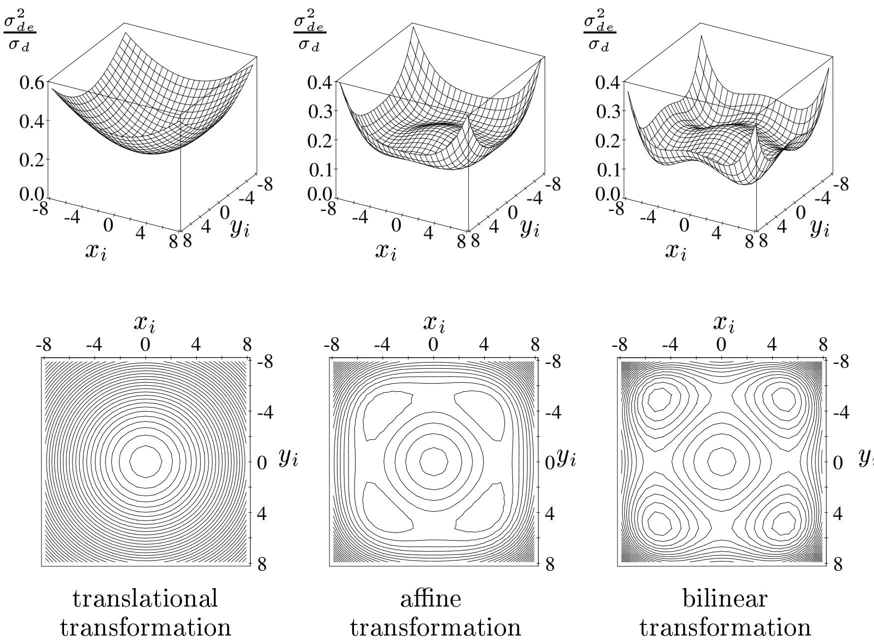


Fig. 2. Spatial disparity error variance distribution inside blocks of the size 17x17 pel. Visualised as 3-D diagram and height contours for three different transformations (other scale for translational transformation).

the 3-D diagrams of the spatial error variance distribution, the height contours are shown. There, the position of the minimum can be recognised more easily.

The first remarkable result is that the error variance in the central postition is exactly the same for all three transformations. This means that the estimation error variance can not be decreased by using more sophisticated transformations, if the transformed image block is concentrically arranged arround the evaluated pel. This seems to contradict to previous investigations which showed that affine transformation is superior to translational even in case of concentrically arranged blocks [7]. Of course, this derivation is only valid, when the real desparity field can be modeled using an AR(1)-process. This applies to natural objects with curved surfaces. In case of large planar surfaces which are arbitrarilly oriented in 3-D space, the model of an AR(1) process is not appropriate. Thus, affine or bilinear estimation can provide more accurate estimates than translational transformation even if the reference block is concentrically arranged arround the evaluated pel.

The second result is that there are four eccentric minima in case of affine and bilinear transformation, which are eccentrically arranged arround the center of the block. An investigation with varying autocorrelation coefficients shows that their position (x_{min}, y_{min}) is proportional to the chosen block size N and varies slightly with ρ_{dd}. It varies less than $0.03N$ for $0.75 < \rho_{dd} < 1$, the typical range of autocorrelation coefficients of real disparity fields. In case of a block size of 17x17 pel (N=17) the position of the minimum is $x_{min} = y_{min} = 4$ for affine and $x_{min} = y_{min} = 5$ for bilinear transformation.

The third result is that ρ_{dd} effects the error variance itself. The error variance rises with decreasing autocorrelation ρ_{dd} between neighbouring disparities. The reason is that with small statistcal bindings between neighbouring disparities the implicit assumption about deterministic course of the disparity that results from the 2-D transformation involves larger model failures.

The fourth result is that ρ_{dd} also effects the ratio between the error variance in the center $\sigma^2_{de,cen}$ and the error variance in the eccentric minimum $\sigma^2_{de,ecc}$. The ratio $g_{ecc} = \sigma^2_{de,cen}/\sigma^2_{de,ecc}$ is denoted as the gain, since an according gain is expected from an estimator that arranges the reference block eccentrically instead of concentrically. In case of affine transformation the gain varies between 1.2 and 1.35. In case of bilinear transformation a gain between 1.45 and 1.95 can be archieved. Since the gain is higher in case of bilinear transformation it is applied to the new estimator.

4 Disparity estimation with eccentric image blocks

As shown in Fig. 2, the error distribution inside the image blocks shows four eccentric minima. This knowledge shall now be exploited for an improvement of the estimator. In all known approaches [4] [7] [10] [12], the image blocks are arranged concentrically arround the evaluated pel, where the estimation error variance is not minimal according to the analytically derived error distribution.

A new estimator is therefore proposed that arranges the image blocks eccentrically arround the evaluated pel. Since four minima exist, there are four possible arrangements of the image block such that the evaluated pel lies in one of the minima (Fig. 3).

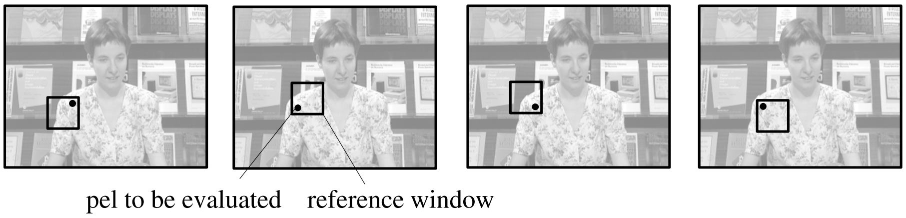


Fig. 3. Four possible eccentric arrangements of the reference block around the pel to be evaluated

With the eccentric arrangement of the image blocks, a gain in terms of reducing the estimation error variance is obtained. This gain depends on the ratio between the error variance in the center $\sigma^2_{de,cen}$ and the error in the minima $\sigma^2_{de,ecc}$ and is expressed by $g_{ecc} = \sigma^2_{de,cen}/\sigma^2_{de,ecc}$.

A further improvement is expected by estimating with all four possible arrangements of the image block. Thus, the proposed estimator averages the four possible estimates arithmetically (15).

$$d_{res} = \frac{1}{4}\left(d_1 + d_2 + d_3 + d_4\right) \tag{15}$$

Assuming that the error involved by each observed pel in the image block is uncorrelated and has the same statistical properties, the estimation error variance decreases reciprocally with the number of observed pels when using an Maximum-Likelihood (ML) estimator. Since the four eccentric image blocks partly overlap, the number of observed pels is not increased by a factor of 4. The computation of the number of observed pels is derived in Fig. 4 results in 2.64 times more observed pels. Thus a maximum gain due to averaging of $g_{av} < 2.64$ is expected from the new estimator.

The same gain can not be expected from a single estimation with a block size b_4, because a single estimation with larger blocks leads to additional model failures that decrease the accuracy of the estimator. In case of the new estimator the block size of each individual estimation is not changed and thus no additional model failures are caused.

Beside the possibility of arithmetically averaging the four estimates, also a more appropriate averaging might be used that leads to more robust estimates. Such improvements are envisaged as a next step, but have not been investigated so far.

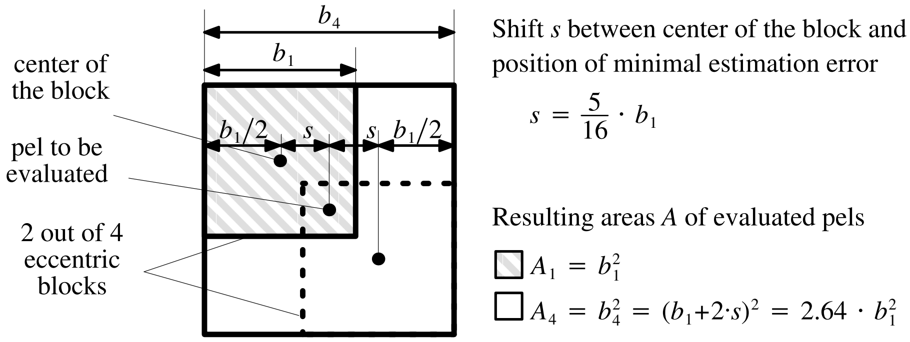
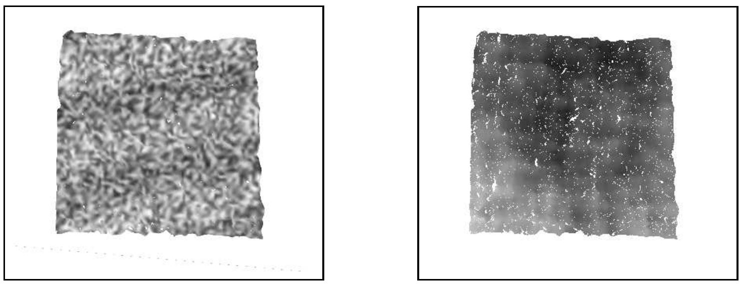


Fig. 4. Relation between the number of observation points in a single block of size b_1 and four partly overlapping blocks of size b1 which cover the same area as a single block of size b_4

5 Experimental results

In this section the theoretically derived error variance distribution and the gain obtained by the proposed estimator are experimentally verified. For a numerical verification of the error variance distribution and the derived gain of the new estimator, the real disparity of the evaluated image pair has to be known. Therefore, two synthetic image pairs were created by rendering two virtual 3-D objects. In order to consider the camera noise, an additional Gaussian noise was added to the synthetic images. Since the real disparity is known, the disparity estimation error can be measured.

In Fig. 5 the left luminance image and the real disparity field of a synthetic image pair called SynAR is shown.

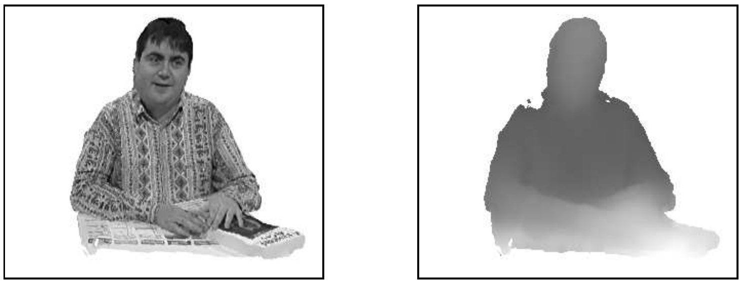


a) synthesised luminance image b) real disparity field

Fig. 5. Synthetic image SynAR

A lowpass filtered random noise is chosen for the texture of the SynAR object. The surface of the object approximates a depth map which results from a stationary AR(1) process. Thus the disparity estimation results should confirm the analytically derived disparity estimation error variance distribution. The real disparity field has an autocorrelation coefficient of $\rho_{dd} = 0.995$.

The second image pair was created from a virtual 3-D object that is reconstructed from a real video communication scene. It is called SynLudo and is shown in Fig. 6. The real disparity field of SynLudo shows discontinuities and is not stationary. Therefore the AR(1)-process can not approximate it as accurately as in case of SynAR. The real disparity field has an autocorrelation coefficient of $\rho_{dd} = 0.98$.



a) synthesised luminance image b) real disparity field

Fig. 6. Synthetic image SynLudo

A block size of 17x17 pel was chosen for the experimental verification. In table 2 the estimation errors variances of the different estimators are shown.

Table 2. Measured estimation error variances of the different estimators

Estimation		SynAR	SynLudo
One concentrically arranged block:	$\sigma^2_{de,cen}$	0.4803 pel^2	0.1214 pel^2
One eccentrically arranged block:	$\sigma^2_{de,ecc}$	0.3096 pel^2	0.0901 pel^2
Four averaged, eccentrically arranged blocks:	$\sigma^2_{de,av}$	0.2070 pel^2	0.0698 pel^2

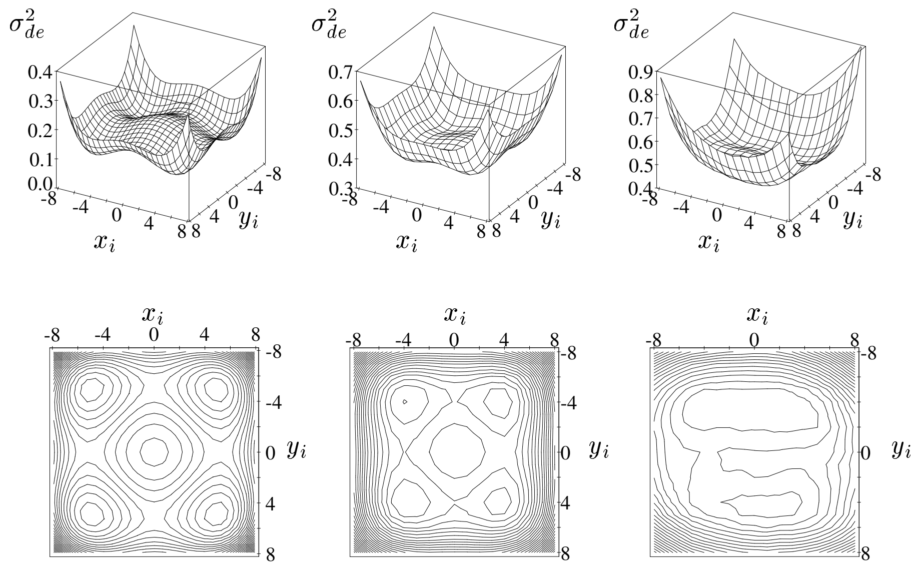
The gain of the proposed estimator is computed from the measured error variances. In case of SynAR the following gain is optained:

$$g_{ecc} = \frac{\sigma^2_{de,cen}}{\sigma^2_{de,ecc}} = \frac{0.4803}{0.3096} = 1.5514 \quad , \quad g_{av} = \frac{\sigma^2_{de,ecc}}{\sigma^2_{de,av}} = \frac{0.3096}{0.2070} = 1.4967$$

In case of SynLudo the gain is

$$g_{ecc} = \frac{\sigma^2_{de,cen}}{\sigma^2_{de,ecc}} = \frac{0.1214}{0.0901} = 1.3474 \quad , \quad g_{av} = \frac{\sigma^2_{de,ecc}}{\sigma^2_{de,av}} = \frac{0.0901}{0.0698} = 1.2908$$

Fig. 7 shows the comparison between the theoretically derived and experimentally found disparity estimation error variance distribution.



theoretically derived SynAR SynLudo

Fig. 7. Comparison between the theoretically derived disparity estimation error variance distribution inside the image blocks and the experimentally found one. Visualised as 3-D diagram and height contours.

The theoretically derived and the measured error distributions are quite similar. The measured position of the minima almost fits to the analytically derived one. But there is a difference between the expected error variance ratio of $g_{ecc} = 1.93$ and the measured one of $g_{ecc} = 1.55$ in case of SynAR. This difference is caused by restricted accuracy of the virtuel 3-D object and error during the image synthesis. On the one hand, the object surface is represented by a triangular mesh causing piecewise linear surfaces that differ from an ideal AR(1) process. On the other hand the applied image synthesis approach just uses bilinear texture interpolation and does no anti-aliasing filtering. In case of SynLudo a gain of $g_{ecc} = 1.35$ is achieved.

¿From averaging, a maximum gain of $g_{av} = 2.64$ can be expected. It turns out that only a gain of $g_{av} < 1.5$ is achieved during the experimental investigations. This indicates that the error between neighbouring pels is correlated. The total gain of the presented estimator is therefore $g = g_{ecc} \cdot g_{av} = 1.55 \cdot 1.5 = 2.33$.

With real images, a subjective assessment of the estimates is carried out by visualising the 3-D cloud of points which results from a triangulation of corresponding image points. It turns out that the robustness is obviously improved due to averaging the four estimates using a simple outlier detection and removal that evaluates the empirical variance of the four estimates.

6 Conclusions

A block-based disparity estimator is proposed in this paper that considers the non-uniform spatial distribution of the disparity estimation error inside an image block.

The presented approach is based on Buschmann's analytical derivation of the spatial distribution of the displacement estimation error inside image blocks which are mapped into another image plane using a 2-D transformation [13]. By restricting the image block transformations to horizontal epipolar geometry, the horizontal component of the displacements is turning into disparity and the vertical displacement is equal to zero. Since Buschmann assumed statistically independent displacement components, also the estimation error variance distribution of the horizontal displacement component corresponds to the one of the disparity estimate.

For the derivation of the estimation error variance, the stochastic process for modelling displacement vector fields [13] is adapted to model disparity fields. The zero-mean AR(1) process is extended by an additional mean value which models the offset in disparity fields that origins from the stereoscopic camera setup. The real disparity field is thus described by its autocorrelation coefficient ρ_{dd}, its variance σ_d^2 and its mean value m_d.

The dependency of the spatial distribution of the disparity estimation error variance inside the reference block on the parameters of the AR(1) process is derived in a next step. It turns out that m_d does not influence the distribution and that σ_d^2 scales the error variance linearly. The error distribution non-linearly depends on ρ_{dd}. Comparing the three different investigated transformations, it is shown that the estimation error variance in the center of the blocks is identical for translational, affine and bilinear transformations. This observation seems to contradict to previous investigations which showed that affine transformation is superior to translational transformation in case of almost planar surfaces and an estimation technique that uses concentrically arranged image blocks [7]. The reason for this contradiction is that the disparity field of planar surfaces, having a deterministic linear course, can not be modeled by an AR(1) process. An AR(1) process rather models disparity fields caused by more complex surfaces.

The position of the four minima is analytically determined for affine and bilinear transformation. It is proportional to the chosen block size N and varies slightly with ρ_{dd}. It varies less than $0,03N$ for $0.75 < \rho_{dd} < 1$. In case of bilinear transformation the position is around $(\pm 5, \pm 5)$ and in case of affine transformation around $(\pm 4, \pm 4)$ with respect to a local coordinate system in the center of the block and a block size of 17x17 pel.

Based on the evaluation of the estimation error distribution, a new estimator is proposed. It arranges the image block eccentrically around the evaluated pel such that it lies in one of the minima of the estimation error variance distribution.

Since the position of the minima is almost constant the arrangement does not depend on the statistic properties of the disparity field.

The gain of the proposed estimator compared to conventional estimation with a concentrically arranged block is expressed by a gain factor g. The gain obtained by eccentrically arranging the blocks is expressed by the ratio between the error variance in the minimum $\sigma^2_{de,ecc}$ and the error variance in the center of the block $\sigma^2_{de,cen}$, leading to $g_{ecc} = \sigma^2_{de,cen}/\sigma^2_{de,ecc}$. This gain non-linearly depends on ρ_{dd}. For common autocorrelation values around $\rho_{dd} = 0.995$, a gain around $g_{ecc} = 1.93$ is obtained in case of bilinear transformation and a gain around $g_{ecc} = 1.83$ is achieved in case of affine transformation. Since the error variance is smaller in case of bilinear transformation, bilinear transformation is applied in the proposed estimator.

The estimation error variance is further decreased by arithmetically averaging the four possible estimates. Assuming that the error of each observed pel is uncorrelated, the error variance decreases reciprocally with the total number of pels in the reference blocks when using a Maximum-Likelihood estimator. Because the eccentric blocks partly overlap, the number of pels increases by a factor of 2,64. Accordingly, a maximum additional gain of $g = 2.64$ can be obtained by averaging the four estimates. The same gain can not be obtained by using a single block with the same number of pels, since this would cause additional model failures.

The proposed estimator is tested in order to verify the analytically derived gain using synthetic and real images and a constant block size of $N = 17$ pel. Synthetic images are created by rendering two virtual 3-D objects: one with a surface whose depth is created using a AR(1) process, called SynAR, and another object that was reconstructed from a videocommunication scene, called SynLudo. The disparity fields of both objects can not be modelled correctly by the investigated transformations.

The analytically determined and measured error distribution are quite similarity. The measured position of the minima almost fits to the analytically derived one. But there is a difference between the expected error variance ratio of $g_{ecc} = 1.93$ and the measured one of $g_{ecc} = 1.55$ in case of SynAR. This difference is caused by restricted accuracy of the virtuel 3-D object and error during the image synthesis. On the one hand, the object surface is represented by a triangular mesh causing piecewise linear surfaces that differ from an ideal AR(1) process. On the other hand the applied image synthesis approach just uses bilinear texture interpolation and does no anti-aliasing filtering. In case of SynLudo a gain of $g_{ecc} = 1.35$ is achieved.

For a further reduction of the estimation error due to averaging, a maximum value of $g_{av} = 2.64$ can be expected. It turns out that only a gain of $g_{av} = 1.5$ is achieved. This indicates that the error between neighbouring pels is correlated. The total gain of the presented estimator is therefore $g = g_{ecc} \cdot g_{av} = 2.33$.

With real images a subjective assessment of the estimates is carried out. It turns out that the robustness is obviously improved due to averaging the four estimates using a simple outlier detection and removal that evaluates the empirical variance of the four estimates. Further improvements are expected but not yet verified by a more appropriate averaging of the estimates, considering disparity discontinuities and individual disparity estimation errors.

References

1. M. J. P. M. Lemmens, "A survey on stereo matching techniques", International Archives of Photogrammetry and Remote Sensing, Vol. 27, Comm. V., pp. 11-23, 1988.
2. J. K. Aggarwal, N. Nandhakumar, "On the computation of Motion from Sequences of Images - A Review", Proceedings of the IEEE, Vol. 76, No. 8, August 1988.
3. U. R. Dhond, J. K. Aggarwal, "Structure from Stereo - A Review", IEEE Trans. on Systems, Man, and Cybernetics, Vol.19, No. 6, Nov./Dec. 1989.
4. H.G. Musmann, P. Pirsch, H.J. Grallert, "Advances in picture coding", Proceedings of the IEEE, Vol. 73, No. 4, S. 523-548, April 1985
5. M. Bierling, "Displacement estimation by hierarchical block matching", Proc. 3rd SPIE Symposium on Visual Communications and Image Processing, Cambridge, USA, pp. 942-951, November 1988
6. J.W. Kim, S.U. Lee,"On the hierarchical variable block size motion estimation technique for motion sequence coding", Proc. SPIE Visual Communications and Image Processing '93, Vol. 2094-I, pp. 372-383, Cambridge, MA, USA, November 1993
7. W. Foerstner, "Quality assessment of object location and point transfer using digital image correlation techniques", Int. Arch. of Photogrammetry and Remote Sensing, Vol. 25, Part A3a, C3, 1984.
8. C. Fuh and P. Maragos, "Motion displacement estimation using an affine model image matching", Optical Engineering, Vol. 30, No. 7, pp. 881-887, July 1991
9. H. Li and R. Forchheimer, "A new motion compensation technique for video compression", Proc. IEEE Int. Conf. Acoust., Speech. Signal Processing, Minneapolis, MN, USA, pp. 441-444, May 1993
10. A. W. Gruen, E. Baltsavias, "Geometrically Constrained Multiphoto Matching", Photogrammetric Engineering and Remot Sensing, Vol. 54, No. 5, May 1988, pp 633-641.
11. Y. Xiong, S. Shafer, " Hypergeometric Filters for Optical Flow and Affine Matching", Int. Journal of Computer Vision, Vol. 24, No. 2 1997, pp 163-177.
12. V. Seferidis, M. Ghanbari, "Generalized block matching motion estimation", SPIE Symposium on Visual Communications and Image Processing, Boston, USA, Vol. 1818, S. 110-118, November 1992.
13. R. Buschmann, "Analytische Bestimmung der Schätzfehlervarianzen von Displacementschätzverfahren der Bewegtbildcodierung", Ph. D Thesis, Hannover, Germany, 1997.
14. R. Buschmann, "Efficiency of displacement estimation techniques", Image Communication, 10/1997, pp. 43-61.
15. R. Koch, "Automatische Oberflächenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Rundum-Ansichten", Dissertation, Universität Hannover, 1996
16. D. E. Wells, E. J. Krakiwsky,"Method of least squares", Dept. of Surveying Eng. Reports, University of New Brunswick, Frederikton, Canada, 1971.
17. L. Falkenhagen, "Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints", International workshop on SNHC and 3D Imaging, September 5-9, 1997, Rhodes, Greece.