Robust Concealment for Erroneous Block Bursts in Stereoscopic Images

Sebastian Knorr, Carsten Clemens, Matthias Kunter and Thomas Sikora
Department of Communication Systems
Technical University of Berlin, 10587 Berlin, Germany
{knorr|clemens|kunter|sikora}@nue.tu-berlin.de

Abstract

With the increasing number of image communication applications especially in the low complexity domain, error concealment has become a very important field of research. Since many compression standards for images and videos are block-based a lot of methods were applied to conceal block losses in monocular images. The fast progress of capture, representation and display technologies for 3D image data advances the efforts on 3D concealment strategies. Because of their psycho-visual characteristics, stereoscopic images have to fulfill a very high quality demand. We propose an algorithm that makes use of the redundancies between two views of a stereo image pair. In many cases erroneous block bursts occur and can be highly disturbing, thus we will mainly concentrate on these errors. In addition, we focused on the quality assessment of several error concealment strategies. Beside the objective evaluation measures, we carried out a subjective quality test following the DSCQS methodology as proposed by MPEG. The results of this test demonstrate the efficiency of our approach.

1. Introduction

For transmission of stereoscopic images over band restricted channels monoscopic source coders are often used, i.e. both views of a stereo image pair are coded independently. Hence redundancies between the two images are not eliminated and can be utilized for error concealment in the decoding process.

In many international standards block-based transform together with variable length coding is applied for source coding (e.g. [7]). Due to variable length coding, a single bit error can cause synchronization loss between the en- and decoder which leads to burst errors until the next synchronization mark.

The perception of depth is highly affected by lost blocks in one channel. While in a monoscopic scenario interpolation techniques yield satisfactory error concealment results,
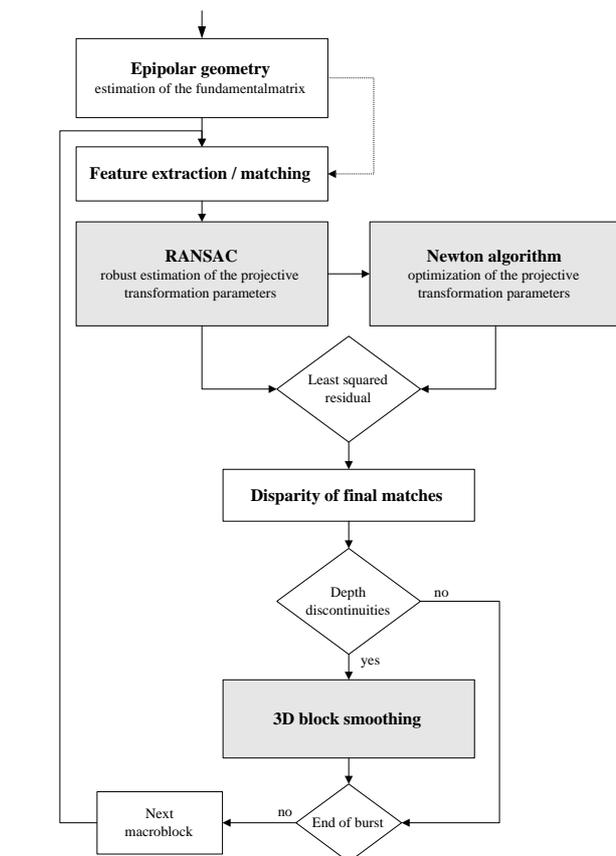

Figure 1. Flowchart of the proposed algorithm

they are not sufficient for stereoscopic image pairs [5], since the information of depth is not preserved. We will show in a subjective evaluation procedure that even small image degradations in one view tend to result in noticeable perceptional disruptions. To our knowledge, only a little work in the field of error concealment for stereoscopic images has been published [1].

In our approach we handle every macro-block of an er-

ror burst separately. Thus, the applied algorithms are introduced for the simple case of single block loss and enhancements, if needed, are discussed. The proposed method locates the corresponding block in the undisturbed image and aligns it to the different perspective of the damaged image. Fig. 1 shows the flowchart of our algorithm.

Without loss of generality we assume a block loss in the left image of a stereo image pair (fig. 2). The alignment is realized by a *projective transformation* $\mathcal{T}$ [5], which is represented by the vector of *transformation parameters* $\mathbf{k}$:

$$
\begin{aligned}
(x_r, y_r) &= \mathcal{T}(\mathbf{k}; (x_l, y_l)) \\
\mathbf{k} &= [a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2]^T \\
x_r &= \frac{a_1 x_l + a_2 y_l + a_3}{c_1 x_l + c_2 y_l + 1} \\
y_r &= \frac{b_1 x_l + b_2 y_l + b_3}{c_1 x_l + c_2 y_l + 1}
\end{aligned} \tag{1}
$$

First, corresponding pixel pairs (matches) around the erroneous region are identified using *feature matching* and the principles of epipolar geometry (section 2). To reduce the negative effect of outliers, i.e. matches which are badly localized or even false, we use a robust estimation to set an initial guess of the projective transformation parameters (section 3). We will show the advantages of the random sample consensus (RANSAC) versus the M-estimation technique.

Depending on the texture of the lost block, a single pixel shift can have a significant effect on the visual perception. Therefore the transformation parameters $\mathbf{k}$ are optimized with respect to the surrounding pixels of the lost block by the *Newton Method*, which is described in section 4. In case of depth discontinuities even the optimized projective transformation can cause distortions which could negatively influence the stereoscopic depth perception. A block smoothing could lessen this effect. We refer to this method in section 5. In section 6 objective simulation results and the subjective evaluation results of a psycho-visual test using the *Double Stimulus Continuous Quality Scale* (DSCQS) method are presented. Section 7 concludes the paper.

2. Feature Extraction/Matching

2.1. Feature Point Detection

Prior to the matching algorithm some so called *feature points* surrounding the lost block have to be determined. These points are selected by the Harris corner detector [3], which is based on gradient measurement. Additionally, the distance to the lost block and the number of *feature points* within a certain range to each other are taken into account, to select the most desirable *feature points*.

2.2. Feature Matching

Matching *feature points* in different images is probably the most difficult problem. To find the appropriate correspondences of the *feature points* in the right image, we utilize the epipolar geometry between both views. The search range can be reduced from the 2D image plane to a line, the so called epipolar line. Knowing the fundamental matrix $\mathbf{F}$, i.e. the algebraic representation of the epipolar geometry [4], the projective mapping from points to lines is given by the following equation:

$$
\mathbf{l_r^i} = \mathbf{F}\mathbf{x_l^i} \qquad i = 1 \ldots n \tag{2}
$$

$\mathbf{F}$ is the fundamental matrix and $\mathbf{l_r^i}$ is the epipolar line in the right image corresponding to the *feature point* $\mathbf{x_l^i}$ in the erroneous left image. As matching score the normalized cross correlation is used. Figure 2 demonstrates the matching of feature points surrounding a lost block. To reduce the
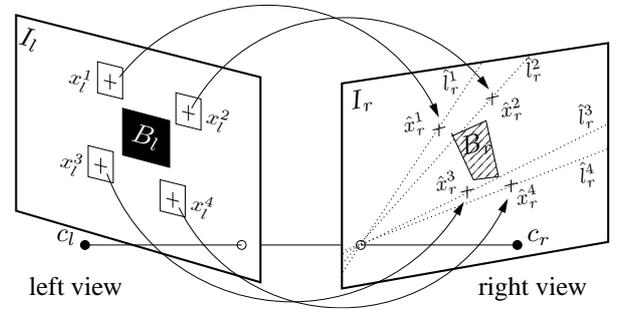

Figure 2. Corresponding blocks in stereo images

probability of mismatches, we back-project the corresponding *feature points* from the right image to epipolar lines in the erroneous left image with the inverse fundamental matrix. The Euclidean distance between the *feature points* and the epipolar lines of the correspondences is a measure of goodness. For further computations we use only the best matches, i.e. matches with small distances to their epipolar lines and with a correlation score higher than a predefined threshold.

3. Robust Initialization of Transformation Parameters

This section describes the initialization phase for the estimation of the transformation parameters. Though eight parameters have to be estimated for a projective transformation model, only four correspondences with a total of eight equations would be sufficient, since none of the matches is

a mismatch. Thus, we increase the number of matches for further computations. To reduce the effects of outliers, several robust optimization methods are discussed in the literature (e.g. [2]). We compare the M-estimator and the random sampling consensus (RANSAC) concerning the robustness and quality of estimating good initial values of the transformation parameters. Section 6 will show some results.

3.1. M-Estimation

The M-estimator is a robust regression method to reduce the effects of outliers by minimizing the following *error function*:

$$\min_{\mathbf{k}} \sum_i \rho(r_i), \qquad (3)$$

where r_i is the residual of the ith datum, i.e. the Euclidian distance of the observation and its fitted location and ρ is the *Tuckey function*:

$$\rho(r_i) = \begin{cases} \frac{(c\sigma)^2}{6} \left(1 - \left[1 - \left(\frac{r_i}{c\sigma} \right)^2 \right]^3 \right) & if |r_i| \leq c\sigma \\ \frac{(c\sigma)^2}{6} & otherwise, \end{cases} \qquad (4)$$

σ is the estimated standard deviation defined as

$$\sigma = 1,4826 \; [1 + 5/(Q-P)] \, median \, |r_i| \qquad (5)$$

and c is a tuning constant. A test series yields best results using $c \approx 5$. Q is the number of equations obtained from the matches. To minimize (3) the derivative $\Psi(x) = \frac{d\rho(x)}{dx}$ has to be used. $\Psi(x)$ is also called the *influence function*. The M-estimator of $\mathbf{k}$ is a solution of $P=8$ equations

$$\sum_i w(r_i) \, r_i \, \frac{\partial r_i}{\partial k_j} \quad for \, j = 1 \ldots P, \qquad (6)$$

where $w(x) = \frac{\Psi(x)}{x}$ is a *weight function*:

$$w(r_i) = \begin{cases} \left[1 - \left(\frac{r_i}{c\sigma} \right)^2 \right]^2 & if |r_i| \leq c\sigma \\ 0 & otherwise, \end{cases} \qquad (7)$$

Because $w(r_i)$ is a non-linear function, iterative numerical methods have to be used to solve (7).

3.2. RANSAC

In this subsection we outline the random sampling consensus (RANSAC), which is essentially a Monte Carlo type approximation algorithm. The basic approach of this optimization method is quite different to the M-estimator. Rather than using all of the matches (inliers and weaker weighted outliers), RANSAC uses only a set of sub-samples

to find the approximately optimal solution with high probability. To calculate the transformation parameters, each sub-sample consists of four feature matches. The sufficient number of sub-samples can be determined with the following equation:

$$N_{sub} = \frac{\log(1-p)}{\log(1 - (1-\epsilon)^4)}, \qquad (8)$$

where p is the probability to get at least one sub-sample which consists of only good correspondences, and ϵ is the assumed percentage of outliers. One major benefit of the RANSAC algorithm is the ability to cope with a large proportion of outliers.

As a criterion for the goodness of the transformation parameters we define the sum of squared residues:

$$R(\mathbf{k}) = \sum_{(x_l, x_r) \in b} [(I_r(x_r, y_r) - I_l(x_l, y_l)) \cdot w]^2 \stackrel{!}{=} min. \qquad (9)$$

(x_l, y_l) are the positions of the pixels within the boundary region b in the left image and (x_r, y_r) are the corresponding positions in the right image according to (1). w is a weighting coefficient. The probability of discontinuities in depth arises with the distance to the erroneous block. Therefore, to minimize the influence of this squared residues, w decreases with the distance. The transformation parameters $\mathbf{k}$ which yield the best results, i.e. the minimum of (9), are taken as initial values for the Newton method.

4. Newton Method

In this section we describe the optimization of the transformation parameters $\mathbf{k}$. We minimize a *cost function* $C(\mathbf{k})$ which depicts the sum of quadratic errors between *boundary pixels* placed around the lost block in the left image I_l and its corresponding pixels in the right image I_r with respect to $\mathbf{k}$. *Boundary pixels* indicate a L pixel wide region around the block. In case of error bursts, only undisturbed regions are used. The cost function is represented by

$$C(\mathbf{k}) = \frac{1}{2} \sum_{(x_l, x_r) \in b} [I_r(x_r, y_r) - I_l(x_l, y_l)]^2. \qquad (10)$$

Assuming the boundary region contains M pixels we can build two vectors $\mathbf{p_{bl}} \in \mathbb{R}^M$ and $\mathbf{p_{br}} \in \mathbb{R}^M$. Eq. (10) becomes

$$C(\mathbf{k}) = \frac{1}{2} (\mathbf{p_{br}}(\mathbf{k}) - \mathbf{p_{bl}})^T \cdot (\mathbf{p_{br}}(\mathbf{k}) - \mathbf{p_{bl}}). \qquad (11)$$

The optimum for $\mathbf{k}$ results from the global minimum of $C(\mathbf{k})$:

$$\mathbf{k_{opt}} = \arg \min_{\mathbf{k}} C(\mathbf{k}) \qquad (12)$$

For local minima $\mathrm{grad_k}(C(\mathbf{k}))$ is a zero vector. In this approach we determine the null of the gradient using the Newton method for multidimensional functions [9]. The iterative rule for updating the transformation parameter vector $\mathbf{k}$ is:

$$\mathbf{k}(n+1) = \mathbf{k}(n) - \mathbf{H_k}^{-1}(C(\mathbf{k})) \cdot \mathrm{grad_k}(C(\mathbf{k}))\Big|_{\mathbf{k}(n)} \quad (13)$$

$\mathbf{H_k}^{-1}(C(\mathbf{k}))$ marks the *Hessian* of the cost C with respect to $\mathbf{k}$. The entries $\mathbf{H_k}^{ij}$ of this symmetric matrix in the ith row and the jth column are:

$$\begin{aligned}
\mathbf{H_k}^{ij}(C(\mathbf{k})) &= \left[\frac{\partial^2 \mathbf{p_{bl}}}{\partial k_i\,\partial k_j}\right]^T (\mathbf{p_{br}}(\mathbf{k}) - \mathbf{p_{bl}}) \\
&+ \left[\frac{\partial \mathbf{p_{br}}(\mathbf{k})}{\partial k_i}\right]^T \frac{\partial \mathbf{p_{br}}(\mathbf{k})}{\partial k_j}
\end{aligned} \quad (14)$$

The first addend in (14) is a weighted sum of pixel differences. Since the difference has zero mean, from a minimum number of entries in the border vector $\mathbf{p_{br}}$ this term can be disregarded [9].

The derivatives of vector $\mathbf{p_{br}}$ in the second addend of (14) are defined as:

$$\frac{\partial p_{br}^m}{\partial k_i} = \frac{\mathrm{d}I(x_r, y_r)}{\mathrm{d}x_r}\frac{\partial x_r}{\partial k_i} + \frac{\mathrm{d}I(x_r, y_r)}{\mathrm{d}y_r}\frac{\partial y_r}{\partial k_i}\Bigg|_{(x_{br}^m, y_{br}^m)} \quad (15)$$

p_{br}^m is the mth entry of vector $\mathbf{p_{br}}(\mathbf{k})$ and (x_{br}^m, y_{br}^m) depicts the corresponding pixel position in the right image ($m = 1, \ldots, M$). The spatial derivations of image I_r are computed by a combination of lowpass and derivation filter as introduced by Simoncelli [8]. The filter coefficients are illustrated in table 1.

Table 1. Filter coefficients for spatial derivations [8]

Filter	Lowpass	Derivation
$h(-2)$	0.004504187	-0.108144
$h(-1)$	0.243908	-0.269869
$h(0)$	0.4221	0.0
$h(1)$	0.243908	0.269869
$h(2)$	0.004504187	0.108144

As an initial guess $\mathbf{k}(0)$ for the Newton rule we use the result obtained from the robust estimation using the RANSAC-algorithm. Since the Newton gradient method only reaches local minima we need to have a special focus on this initial parameter set. Figure 3 displays the normalized *cost function* for a lost 16×16 block over the horizontal and vertical translation parameter. Since the cost $C(\mathbf{k})$ depends on 8 variables we reduced the parameter space for better visualization. Even a deviation of 5 pix-
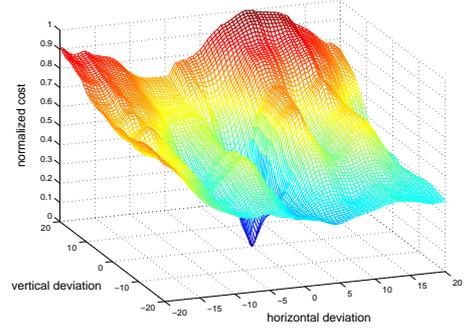

Figure 3. Normalized cost function with respect to horizontal and vertical deviation

els in the initial guess leads to a wrong final solution. Thus, the RANSAC-algorithm described in section 3 is of prime importance.

5. 3D Block Smoothing

Since the projective transformation $\mathcal{T}$ only represents an approximation of a real 3D space model, misalignments between the warped block region $\hat{B}_l$ and its surrounding pixels can occur. Discontinuities in depth are a reliable indication. If the disparities of the matches vary above a given threshold, we apply a smoothing strategy which is closely related to the one introduced by Wang et al. [10] for monoscopic error concealment. Figure 4 shows a scheme for combining boundary pixels of the lost block and the information retrieved from the corresponding picture to generate a smooth block transition. In [1] we introduced this technique as *3D block smoothing* and refer this as 3D-BS.
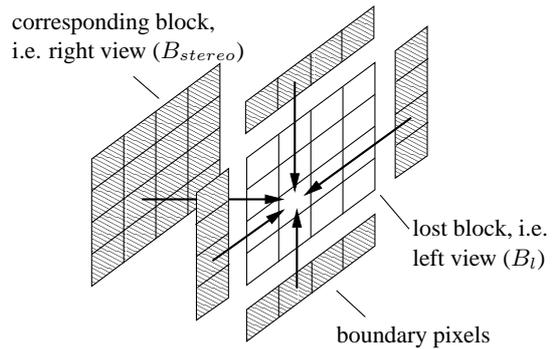

Figure 4. Estimation of lost samples

In Figure 4 B_l represents the lost block of samples (m, n) with values $f_{m,n}$ in the left image. The estimated value of a sample in the lost block is denoted as $\hat{f}_{m,n}$. Pixel

values of the aligned block obtained from the corresponding view are denoted as $f_{m,n}^{stereo}$. We minimize the inter sample variation between neighboring samples within the block B_l and between corresponding samples of B_l and B_{stereo}. The smoothness measure is defined as

$$
\begin{aligned}
\Psi(\hat{f}_{m,n}) &= \frac{1}{2} \sum_{(m,n)\in B_l} \left[w_{m,n}^n \left(\hat{f}_{m,n} - \hat{f}_{m-1,n} \right)^2 \right. \\
&\quad + w_{m,n}^e \left(\hat{f}_{m,n} - \hat{f}_{m,n+1} \right)^2 \\
&\quad + w_{m,n}^s \left(\hat{f}_{m,n} - \hat{f}_{m+1,n} \right)^2 \\
&\quad + w_{m,n}^w \left(\hat{f}_{m,n} - \hat{f}_{m,n-1} \right)^2 \\
&\quad + \left. w_{m,n}^{stereo} \left(\hat{f}_{m,n} - f_{m,n}^{stereo} \right)^2 \right].
\end{aligned}
\tag{16}
$$

Eq. (16) involves samples surrounding the lost block, which will be referred as *boundary pixels* **b**. Pixels from the analogous block can also be considered as *boundary pixels* into a third dimension (Figure 4). The superscripts indicate the direction (north, east, south, west and stereo). The weighting coefficients $w_{m,n}^n$, $w_{m,n}^e$, $w_{m,n}^s$, $w_{m,n}^w$ and $w_{m,n}^{stereo}$ determine the importance of smoothness between a pixel value $\hat{f}_{m,n}$ and its direct neighbors. Using matrix notation, eq. (16) can be written as

$$
\begin{aligned}
\Psi_{ext}(\hat{\mathbf{f}}) &= \frac{1}{2} \left(\|\mathbf{S}_n \hat{\mathbf{f}} - \mathbf{b}_n\|^2 + \|\mathbf{S}_e \hat{\mathbf{f}} - \mathbf{b}_e\|^2 + \right. \\
&\quad + \|\mathbf{S}_s \hat{\mathbf{f}} - \mathbf{b}_s\|^2 + \|\mathbf{S}_w \hat{\mathbf{f}} - \mathbf{b}_w\|^2 + \\
&\quad + \left. \|\mathbf{S}_{stereo} \hat{\mathbf{f}} - \mathbf{b}_{stereo}\|^2 \right).
\end{aligned}
\tag{17}
$$

The matrices $\mathbf{S}_n$, $\mathbf{S}_e$, $\mathbf{S}_s$, $\mathbf{S}_w$ and $\mathbf{S}_{stereo}$ are composed of the weighting coefficients w of eq. (16). Note that the matrices $\mathbf{S_i}$ are sparse and of size $N^2 \times N^2$ (N defines the block size) while the border vectors are of size $N^2 \times 1$. $\mathbf{b_{stereo}}$ contains N^2 nonzero values compared to N nonzero values of the other boundary vectors.

The optimal estimation of $\hat{\mathbf{f}}_{opt}$ can be derived using

$$
\frac{\partial \Psi}{\partial \hat{\mathbf{f}}_{opt}} = 0,
\tag{18}
$$

which yields

$$
\hat{\mathbf{f}}_{opt} = \mathbf{S}^{-1} \mathbf{b}.
\tag{19}
$$

Matrix $\mathbf{S}$ and vector $\mathbf{b}$ are composed as follows:

$$
\begin{aligned}
\mathbf{S} &= \mathbf{S}_n^T \mathbf{S}_n + \mathbf{S}_e^T \mathbf{S}_e + \mathbf{S}_s^T \mathbf{S}_s + \mathbf{S}_{stereo}^T \mathbf{S}_{stereo} \\
\mathbf{b} &= \mathbf{S}_n^T \mathbf{b}_n + \mathbf{S}_e^T \mathbf{b}_e + \mathbf{S}_s^T \mathbf{b}_s + \mathbf{S}_{stereo}^T \mathbf{b}_{stereo}
\end{aligned}
$$

Depending on the distance to the relevant border, we choose decreasing weighting coefficients within $\mathbf{S}_{n,e,s,w}$, while keeping a constant weight within $\mathbf{S}_{stereo}$. The ratio $w_{m,n}^{n,e,s,w}/w_{m,n}^{stereo}$ for every sample (m,n) defines whether $\hat{\mathbf{f}}_{m,n}$ adapts to the borders in direction *north*, *east*, *south* and *west* or it adapts to border *stereo*.

6. Simulation Results

In this section objective and subjective tests are carried out. As an objective quality factor we choose the PSNR of a concealed burst. A psycho-visual test shows the subjective impact of different concealment strategies on test persons. We use the *Double Stimulus Continuous Quality Scale Method* (DSCQS) [6] described in subsection 6.2. Three example image pairs are taken. Example "Hall" and example "Castle" are shown in figure 5 and figure 8. Example "Tower" is similar with higher luminance difference between left and right view.

6.1. Objective Simulation Results

The simulation results of the robust optimization methods for the initialization phase (M-estimator vs. RANSAC) are presented in table 2. The PSNR of the reconstructed block is averaged over 5955 burst errors of 3 consecutively lost 16×16 blocks ("Castle") and 7600 burst errors of the same size ("Hall"). The RANSAC method yields signifi-

Table 2. Average PSNR: M-estimator vs. RANSAC

PSNR in dB	Hall	Castle
M-estimator	23.79	19.06
RANSAC	32.53	20.02

cantly better results than the M-estimator, especially for example "Hall". Furthermore, the transformation failed eight times using the M-estimator, because pixels from outside the image were warped into the erroneous burst.

In the second simulation processes we compare four different error concealment techniques: a monoscopic error concealment technique proposed in [10], the Newton method described in section 4 where the burst is handled once as a whole (NEWTburst) and block by block (NEWT), and the Newton method (NEWT) combined with the 3D-BS described in section 5. Again the PSNR is used as quality factor. 130 burst errors of 8 consecutively lost 16×16 blocks ("Castle") and 169 burst errors of the same size ("Hall") are taken into account. The simulation results of the concealment methods are presented in table 3.

Table 3. Average PSNR in dB for arbitrary burst losses

	mono	NEWTburst	NEWT	3D-BS
Hall	22.10	27.74	29.44	30.78
Castle	18.06	20.79	21.41	22.46

The table shows that the combination of the Newton method and the 3D-BS outperforms the other error conceal-

ment strategies. As expected the monoscopic method yields by far the worst results. Due to the relative small bursts of 8 consecutively lost 16×16 blocks, the Newton method for bursts as a whole even results in a high PSNR, though it is 2 to 3 dB less than the PSNR of the combined Newton and 3D-BS method.

As mentioned in section 5 the 3D-BS is processed only if the disparity variation of the four selected matches (RANSAC) is higher than a predefined threshold. In example "Hall" the execution rate was only 4.51%. Due to larger discontinuities in depth, the execution rate of 3D-BS for example "Castle" was 17.69%.

Figures 5-7 show the concealment results of image pair "Hall" and illustrate the visual impact of the error concealment strategies. The block-PSNR values are 25.41dB (monoscopic), 33.83dB (NEWTburst), 38.42dB (NEWT) and 38.42dB (3D-BS) for the upper error burst and 17.71dB (monoscopic), 18.67dB (NEWTburst), 27.66dB (NEWT) and 27.66dB (3D-BS) for the lower error burst. For these examples no adaptive block smoothing process did occur, so we obtain the same PSNR value for methods NEWT and 3D-BS.

Figures 8-10 show the concealment results of image pair "Castle". The block-PSNR values are 16.87dB (monoscopic), 20.89dB (NEWTburst), 25.01dB (NEWT) and 24.99dB (3D-BS) for the upper error burst and 16.94dB (monoscopic), 16.86dB (NEWTburst), 19.57dB (NEWT) and 19.75dB (3D-BS) for the lower error burst.

6.2. Subjective Stimulation Results

To strengthen the simulation results from section 6.1, a psycho-visual test (DSCQS) with fifteen test subjects was carried out. Active liquid crystal shutter glasses were used to perceive the stereoscopic depth. The subjects were presented with a series of pairs of stereoscopic images. Each concealed image pair was related twice successively to its error free version (reference). The time slot for each stereoscopic image was 8 seconds and the time between two images was 2 seconds. The persons recorded their assessment of the quality of both images (reference and test) on two continuous graphical scales for each test period. A measurement of length makes the subjective score available, which is within a range of 0 to 10.

In our test condition each stereoscopic image pair had three varying error bursts of the following sizes:

- 18 consecutively lost 16×16 blocks

- 12 consecutively lost 16×16 blocks

- 8 consecutively lost 16×16 blocks
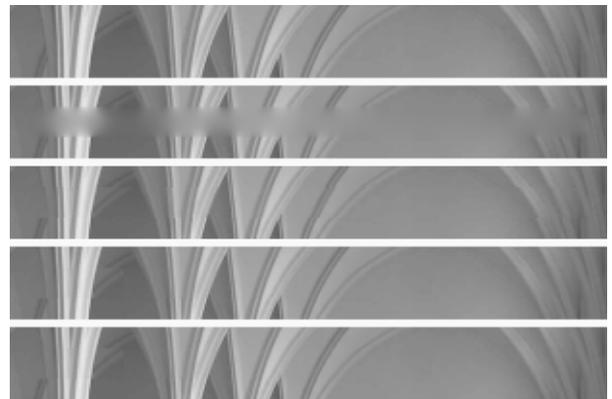

Figure 5. Erroneous left image ("Hall")

Figure 6. Close-up: original (top), mono concealment (2nd), NEWTburst (3rd), NEWT (4th), 3D-BS (bottom)
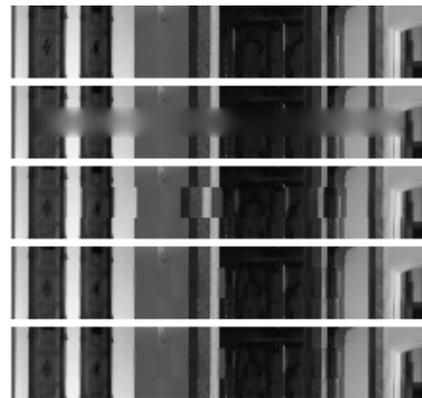

Figure 7. Close-up: original block (top), mono concealment (2nd), NEWTburst (3rd), NEWT (4th), 3D-BS algorithm (bottom)
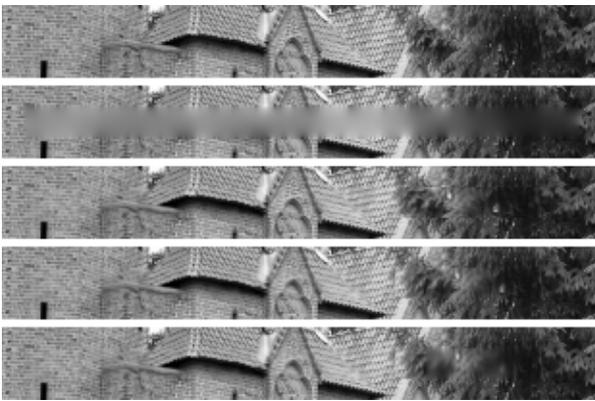
Figure 8. Erroneous left image ("Castle")

Figure 9. Close-up: original block (top), mono concealment (2nd), NEWTburst (3rd), NEWT (4th), 3D-BS algorithm (bottom)

Figure 10. Close-up: original block (top), mono concealment (2nd), NEWTburst (3rd), NEWT (4th), 3D-BS algorithm (bottom)

The evaluation results, i.e. the *mean opinion scores* (MOS), the *difference mean opinion scores* (DMOS) and its standard deviation (S.D.), are presented in table 4.

Table 4. Mean opinion scores

	MOS [test (ref.)]	DMOS (S.D.)
Hall:		
Monoscopic	2.30 (7.96)	5.66 (1.03)
NEWTburst	3.67 (8.04)	4.37 (1.85)
NEWT	7.86 (7.84)	-0.02 (0.64)
3D-BS	7.84 (8.05)	0.21 (0.50)
Castle:		
Monoscopic	2.62 (8.12)	5.50 (1.37)
NEWTburst	4.48 (8.22)	3.74 (2.31)
NEWT	6.58 (8.04)	1.46 (1.75)
3D-BS	6.56 (7.99)	1.43 (1.83)
Tower:		
Monoscopic	2.78 (7.26)	4.48 (1.29)
NEWTburst	5.09 (7.65)	2.56 (2.16)
NEWT	5.43 (7.61)	2.18 (1.76)
3D-BS	5.38 (7.27)	1.89 (1.32)
overall:		
Monoscopic	2.52 (7.88)	5.36 (1.29)
NEWTburst	4.28 (8.03)	3.75 (2.18)
NEWT	6.86 (7.87)	1.01 (1.65)
3D-BS	6.83 (7.87)	1.04 (1.49)

In figure 11 the diagram of the *difference mean opinion scores* (DMOS) for all concealment strategies is displayed. The psycho-visual test shows that our proposed approaches
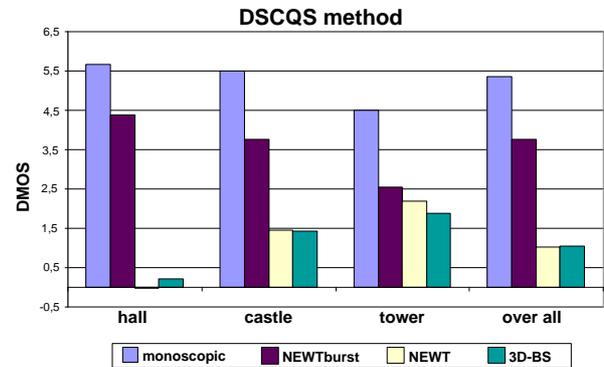

Figure 11. Psycho-visual test results

yield excellent concealment results. The combination of the Newton method and 3D-BS does not result in a higher DMOS than using only the Newton method, although the PSNR was up to 1dB higher. In example "Hall" most of the test subjects could not make a difference between the original and the concealed images using only the Newton method (NEWT). Three test subjects even rated the quality

of the concealed images better than the original.

7. Summary and Conclusion

In this paper we have presented a new method for robust stereoscopic error concealment. Our main attention focused on the concealment of erroneous block bursts. Based on the projective transformation model we applied Newtons algorithm to improve reconstruction results of lost blocks in stereoscopic images. If the transformation model diverges from the 3D structure we additionally applied a block smoothing algorithm to increase the perceived image quality. For the assessment of the proposed method a subjective evaluation procedure was carried out. The main conclusions from the test results are:

(1) The strategy of handling all blocks of consecutive block losses separately outperforms the approach of handling them as a whole using the subjective evaluation criteria. Note that the objective block-PSNR measure does not allow to make such a clear statement.

(2) Smoothing the aligned blocks from the corresponding image towards the border pixels does not really improve the subjective perception quality, although the PSNR measure increases slightly.

(3) In general the PSNR of a recovered erroneous block burst serves as a good rough measure for the quality of perception. Quantitative statements about the performance of different concealment algorithms must be proofed by a subjective quality scale evaluation.

8. Acknowledgment

References

[1] C. Clemens, M. Kunter, S. Knorr, and T. Sikora. A hybrid approach for error concealment in stereoscopic images. *5th International Workshop on Image Analysis for Multimedia Interactive Services*, 2004.

[2] O. Faugeras and Q.-T. Luong. *The Geometry of Multiple Images*. Massachusetts Institute of Technology, 2001.

[3] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the fourth Alvey Vision Conference*, pages 189–192, 1988.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[5] L. M. J. Meesters, W. A. Ijsselstein, and P. J. H. Seuntiens. A survey of perceptual evaluations and requirements of three-dimensional TV. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:381–391, March 2004.

[6] F. Pereira and T. Ebrahimi. *The MPEG-4 Book*. Prentice Hall, 2002.

[7] T. Sikora. MPEG digital video-coding standards. *Signal Processing Magazine, IEEE*, 14(5):82–100, Sep 1997.

[8] E. P. Simoncelli. *Distributed Representation and Analysis of Visual Motion*. MIT Media Laboratory, Cambridge, MA, 1993.

[9] A. Smolic, T. Sikora, and J.-R. Ohm. Long-term global motion estimation and its application for sprite coding, content description, and segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 7:1227–1242, Dec. 1999.

[10] Y. Wang, Q.-F. Zhu, and L. Shaw. Maximally smooth image recovery in transform coding. *IEEE Transactions on Communications*, 1993.