A GRADIENT BASED APPROACH FOR STEREOSCOPIC ERROR CONCEALMENT

Matthias Kunter, Sebastian Knorr, Carsten Clemens, and Thomas Sikora

Department of Communication Systems
Technical University of Berlin, 10587 Berlin, Germany
{kunter|knorr|clemens|sikora}@nue.tu-berlin.de

ABSTRACT

Error concealment is an important field of research in image processing. Many methods were applied to conceal block losses in monocular images. In this paper we present a concealment strategy for block loss in stereoscopic image pairs. Unlike the error concealment techniques used for monocular images, the information of the associated image is utilized , i.e., by means of a projective transformation model, pixel values from the associated stereo image are warped to their corresponding positions in the lost block. The stereoscopic depth perception is much less affected in our approach than using monoscopic error concealment techniques.

1. INTRODUCTION

Common monoscopic source coders can be utilized for stereoscopic image pairs, when the two channels, i.e., left view and right view of the scene, are handled independently. The necessary synchronization of bitstreams can be achieved by multiplexing the bitstream of the two monoscopic coders. In this coding approach redundancies between the channels are not utilized to gain a higher coding efficiency and can therefore be used for error concealment.

In many international standards block-based transform together with variable length coding is applied for source coding (e.g. [1]). 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. In this paper we assume loss of single and consecutive blocks.

The perception of depth is highly affected by a lost block in one channel. While in a monoscopic scenario interpolation techniques yield satisfactory error concealment results, they are not sufficient for stereoscopic image pairs, due to their low pass character. To our knowledge, no specific error concealment strategies for stereoscopic images utilizing the second channel have been proposed.

In our approach, the corresponding block is located in the undisturbed image and aligned to the different perspective of the damaged image. Without loss of generality we assume a block loss in the left image of a stereo image pair (fig. 1). The alignment is realized by a *projective transform* [2].

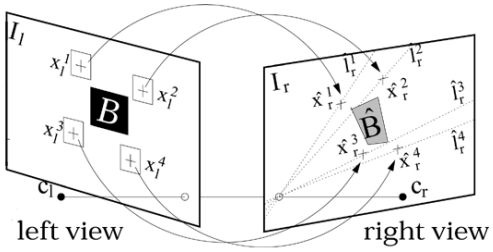

Fig. 1. Corresponding blocks in stereo images

First, the corresponding pixel pairs are identified using the principles of epipolar geometry. In case of occlusion the corresponding pixel can not be found (outliers). To reduce the negative effect of outliers, all matches are weighted by an M-estimator, prior the initial guess of the parameters of the projective transform (section 2).

The projective transformation $\mathcal{T}$ is represented by the vector of *transformation parameters* $\mathbf{k}$:

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

Depending on the texture of the lost block, a single pixel shift can have a significant effect on the psychovisual perception. Therefore the transformation parameters $\mathbf{k}$ are optimized with respect to the surrounding pixels of the lost block by the *Adapted Newton Method*, which is described in section 3. In section 4 simulation results are presented and discussed. Section 5 concludes the paper.

2. POINT CORRESPONDENCES

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 a 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. Correspondences

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:

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

$\mathbf{F}$ 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. To reduce the probability of mismatches, we back-project the matches 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 take only the best matches, i.e. matches with small distances to their epipolar lines and with a correlation score higher than a predefined threshold.

2.3. M-Estimation

The previous section described a robust method of finding point correspondences in a stereo image pair. In this section we estimate the eight parameters of the projective transformation $\mathcal{T}$ from a sufficient number of point correspondences. In our algorithm we use a M-estimator as a robust regression method to reduce the effects of outliers by minimizing the following *error function*:

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

where r_i is the residual of the i^{th} datum, i.e. the Euclidian distance of the observation and its fitted location and ρ is the *Tuckey function* [4]. 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 *N=8* equations

$$\sum_i w(r_i)\, r_i\, \frac{\partial r_i}{\partial k_j} \quad for\ j = 1 \dots N, \tag{4}$$

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

$$w(r_i) = \begin{cases} \left[1 - \left(\frac{r_i}{c\sigma}\right)^2\right]^2 & if |r_i| \le c\sigma \\ 0 & otherwise, \end{cases} \tag{5}$$

σ is the estimated standard deviation defined as

$$\sigma = 1,4826\ [1 + 5/(M - N)]\, median\ |r_i| \tag{6}$$

and c is a tuning constant. A test series yields best results using $c \approx 5$. Because $w(r_i)$ is a non-linear function, iterative numerical methods have to be used to solve (5).

3. ADAPTED NEWTON METHOD

In this section we describe the optimization of the transformation parameters $\mathbf{k}$. This is done by minimizing 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. The cost function is represented by

$$C(\mathbf{k}) = \frac{1}{2} \sum_{(x_l, x_r) \in b} [I_r(x_r, y_r) - I_l(x_l, y_l)]^2, \tag{7}$$

where (x_l, y_l) are the positions of the pixels with in the boundary region b in the left image and (x_r, y_r) are the corresponding positions in the right image according to (1). Assuming the boundary region contains M pixels we can build two vectors $\mathbf{p_{bl}} \in \mathbb{R}^M$ and $\mathbf{p_{br}} \in \mathbb{R}^M$. Eqn. (7) becomes

$$C(\mathbf{k}) = \frac{1}{2}(\mathbf{p_{br}}(\mathbf{k}) - \mathbf{p_{bl}})^T \cdot (\mathbf{p_{br}}(\mathbf{k}) - \mathbf{p_{bl}}). \tag{8}$$

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

$$\mathbf{k_{opt}} = \arg\min_{\mathbf{k}} C(\mathbf{k}) \tag{9}$$

For local minima $\mathrm{grad}_{\mathbf{k}}(C(\mathbf{k}))$ is a zero vector. In this approach we determine the null of the gradient using the Newton method for multidimensional functions [2]. The iterative rule for updating the transformation parameter vector $\mathbf{k}$ is:

$$\mathbf{k}(n+1) = \mathbf{k}(n) - \mathbf{H}_{\mathbf{k}}^{-1}(C(\mathbf{k})) \cdot \mathrm{grad}_{\mathbf{k}}(C(\mathbf{k}))\Big|_{\mathbf{k}(n)} \tag{10}$$

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

$$\mathbf{H}_{\mathbf{k}}^{ij}(C(\mathbf{k})) = \left[\frac{\partial^2 \mathbf{p_{bl}}}{\partial k_i\, \partial k_j}\right]^T (\mathbf{p_{br}}(\mathbf{k}) - \mathbf{p_{bl}}) + \left[\frac{\partial \mathbf{p_{br}}(\mathbf{k})}{\partial k_i}\right]^T \frac{\partial \mathbf{p_{br}}(\mathbf{k})}{\partial k_j} \quad (11)$$

The first addend in (11) 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 [2].

The derivatives of vector $\mathbf{p_{br}}$ in the second addend of (11) are defined as:

$$\frac{\partial p_{br}^m}{\partial k_i} = \frac{\mathrm{d}I(x_r, y_r)}{\mathrm{d}x_r}\frac{\partial x_r}{\partial k_i} + \frac{\mathrm{d}I(x_r, y_r)}{\mathrm{d}y_r}\frac{\partial y_r}{\partial k_i}\Bigg|_{(x_{br}^m, y_{br}^m)} \quad (12)$$

p_{br}^m is the m^{th} 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 an initial guess $\mathbf{k}(0)$ for the Newton rule we use the result obtained from the robust M-estimation. Since the Newton gradient algorithm only reaches local minima we need to have a special focus on this initial parameter set. Fig. 2 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

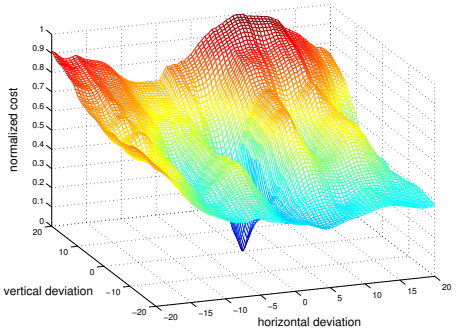

Fig. 2. Normalized cost function with respect to horizontal and vertical deviation in pixels

of 5 pixels in the initial guess leads to a wrong final solution. Therefore, the M-estimator described in section 2 is of prime importance.

For greater robustness we analyzed the convergence of the Newton method with respect to the border pixel size L. The plots in fig. 3 illustrate the normalized *cost function* $C(\mathbf{k})/M$ over the iteration counter for different L. Due to experiments we need a minimum border size about 10 to 15 pixel to reach the minimum at all. The larger the boundary region the faster the algorithm converges. On the other hand, for large L the cost function gets smoother and the minimum can not be found exactly.

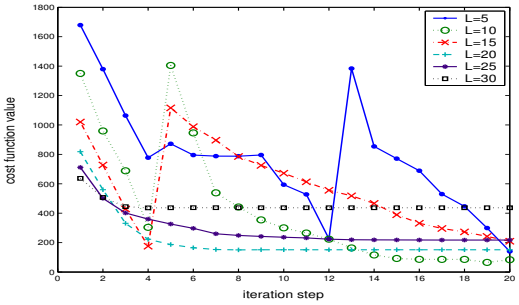

Fig. 3. Convergence of the Newton algorithm for a 16x16 block loss and different pixel border sizes L

This leads to a new strategy for utilizing the Newton method, decreasing successively the border pixel size L after every minimum search. We execute the complete algorithm for decreasing L, where every solution $\mathbf{k}$ is used as initial vector for the the following search. Experiments have shown that this iterative approach outperforms the Newton algorithm with a defined border pixel size.

4. SIMULATION RESULTS

To show the efficiency of our method, we take two example stereo image pairs. Example "Hall" is displayed in fig. 4. Example "Castle" is similar with larger discontinuities in depth. First, we assume several arbitrary block losses in the left images and reconstruct them with three different techniques: a monoscopic error concealment technique proposed in [5], the projective transformation model achieved from the M-estimator described in section 2.3 and the Adapted Newton Method described in section 3. For each reconstructed block we calculated the PSNR. Results were shown in table 1 for block sizes of 8 and 16 pixel.

	Hall		Castle	
PSNR in dB	size=8	size=16	size=8	size=16
Monoscopic	23.16	19.71	17.94	17.26
M-estimator	27.31	24.00	20.92	20.78
Newton	34.28	30.67	22.49	22.95

Table 1. Average block-PSNR for arbitrary block losses

It is obvious that the Adapted Newton Method provides the best results based on PSNR measure. To strengthen these results, a psychovisual evaluation with eight test sub-

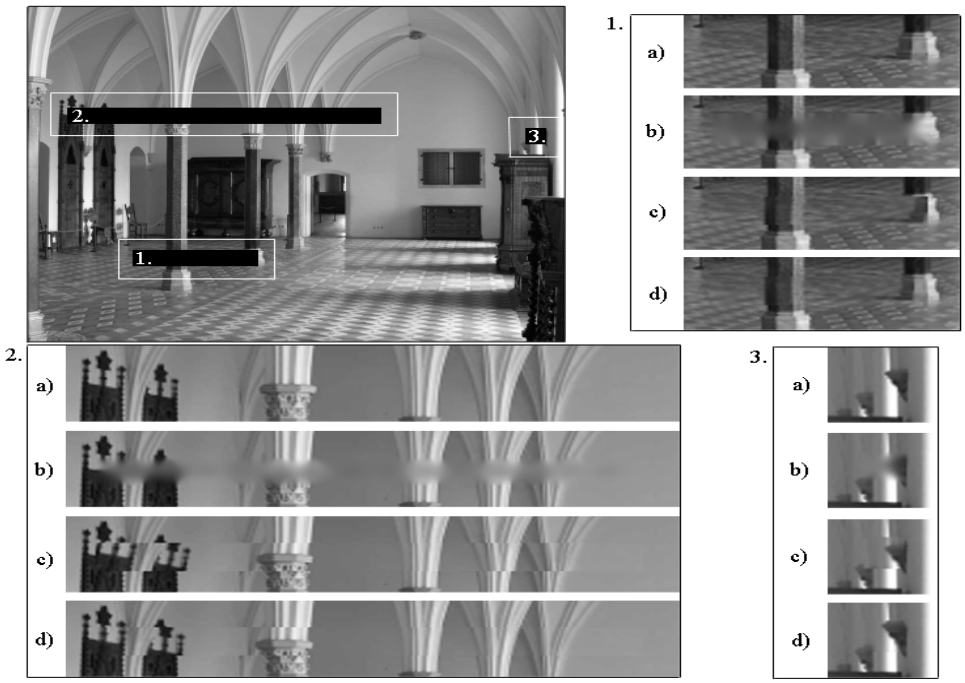

Fig. 4. Different block errors (1.-3.) for image pair "Hall" and restorations: a) original block, b) mono concealment, c) M-estimation, d) Adapted Newton Method

jects showed, that the depth perception is much less affected using the stereo concealment approach, even without the Newton algorithm (i.e. only the M-estimation), than using the monoscopic error concealment technique. Anyhow, the Newton algorithm shows by far the best results, especially when the lost block or block burst is in a plane.

In fig. 4 we show the advantages of our algorithm for three different block errors of image pair "Hall". One major benfit is that bursts of block losses are handled the same way as single block losses. Thus, even for large bursts significant gains on monoscopic concealment strategies are achievable. The block-PSNR values are 22.86 dB (monoscopic) and 26.70 dB (Newton) for the first block burst; 21.72 dB (monoscopic) and 21.76 dB (Newton) for the second block burst, and 18.44 dB (monoscopic) and 21.69 dB (Newton) for the third single block. The reconstruction quality decreases with the presence of great discontinuities in depth because of the 2D transformation constraint.

5. SUMMARY

In this paper we have presented a new approach for stereoscopic error concealment. Based on the projective transformation model, we adapted Newton's algorithm to improve reconstruction results of lost blocks in stereoscopic images. We have shown that the stereoscopic depth perception is much less affected applying our approach than using a monoscopic error concealment technique.

6. ACKNOWLEDGMENT

The work presented was developed within VISNET, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme.

7. REFERENCES

[1] T. Sikora, "Mpeg digital video-coding standards," *Signal Processing Magazine, IEEE*, vol. 14, no. 5, pp. 82–100, Sep 1997.

[2] 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*, vol. 7, pp. 1227–1242, Dec. 1999.

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

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

[5] Yao Wang, Qin-Fan Zhu, and Leonard Shaw, "Maximally smooth image recovery in transform coding," *IEEE Transactions on Communications*, 1993.