A HYBRID APPROACH FOR ERROR CONCEALMENT IN STEREOSCOPIC IMAGES

Carsten Clemens, Matthias Kunter, Sebastian Knorr and Thomas Sikora

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

ABSTRACT

Error concealment for stereoscopic images receives little attention in research of image processing. While many methods have been proposed for monocular images, this paper considers a concealment strategy for block loss in stereoscopic image pairs, utilizing the information of the associated image to fulfill the higher quality demand. We present a hybrid approach, combining a 2D projective transformation and a monoscopic error concealment technique. Pixel values from the associated stereo image are warped to their corresponding positions in the lost block. To reduce discontinuities at the block borders, a monoscopic error concealment algorithm with low-pass properties is integrated. The stereoscopic depth perception is much less affected in our approach than using only monoscopic error concealment techniques.

1. INTRODUCTION

A typical transmission system for monoscopic still or motion images consists of a *source coder*, a *channel coder* and a (error-prone) network. A simple way of coding stereo image pairs is to apply the monoscopic source coder to each channel, multiplex the two bitstreams and perform standard channel coding. In many international standards block-based transform coding is applied for source coding [1], [2]. Because of high decorrelation and energy compaction properties, the *Discrete Cosine Transform (DCT)* is a commonly used unitary and orthogonal transform.

When variable bit length coding is applied, a single bit error can interrupt the synchronization between the encoder and decoder, which leads to information loss until the next synchronization mark. We assume a synchronization mark after each block, although the proposed algorithm can be easily extended to handle consecutive loss of blocks.

Stereoscopic image quality degradation caused by lost blocks is not comparable to the type of degradation caused by quantization or low-pass filtering which was studied in the work of Ijsselstein [3] and Stelmach [4]. The binocular perception of a stereo image pair is highly affected by a lost block in one channel. This is why standard error concealment strategies are not well suited for stereoscopic images. Even reasonably good monoscopic concealment results may have great impairment on the stereoscopic effect. Furthermore extended information, namely the second frame, is not applied. To our knowledge, no specific error concealment strategies for stereoscopic images utilizing the second channel have been proposed so far.

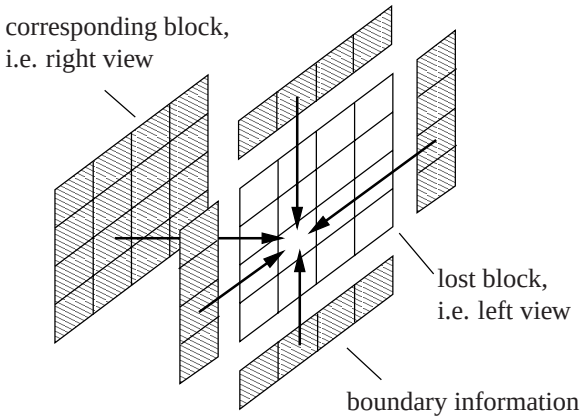

Fig. 1: Estimation of lost samples

Section 2 describes a monoscopic error concealment technique proposed by Wang et al. [5] which yields satisfactory reconstruction results due to a smoothness constraint.

In section 3 this concealment strategy is adapted to the stereoscopic scenario, combining the good perceptual properties (*smoothness*) with additional information of the second frame. Assuming the corresponding block is available, it has to be located and aligned to the different perspective of the damaged frame first. Then, as shown in fig. 1, the boundary information of the lost block can be utilized together with the corresponding sample. Thus, the smoothing towards surrounding samples (border), introduced by Wang et al., is extended by smoothing towards the corresponding block. We refer to this technique as *3D block smoothing* (3D-BS).

In section 4 simulation results are presented and discussed. Section 5 concludes the paper.

2. MONOSCOPIC ERROR CONCEALMENT

The algorithm proposed by Wang et al. [5] handles the estimation of lost coefficients of a unitary block transform. By minimizing the intersample variation within the block and its boundaries this algorithm yields a *maximally smooth* image. A short overview is given below.

Let B represent a block of samples (m,n) with values $f_{m,n}$. Without loss of generality we assume a block loss in the left image of a stereo image pair, denoted by B_l. The estimated value of a sample in the lost block is denoted as $\hat{f}_{m,n}$ for $(m,n) \in B_l$. The smoothness measure proposed by Wang et al. is defined as

$$\begin{aligned}\Psi(\hat{f}_{m,n}) &= \frac{1}{2}\sum_{(m,n)\in B}\Big[w^n_{m,n}\left(\hat{f}_{m,n}-\hat{f}_{m-1,n}\right)^2+\\ &\quad +w^e_{m,n}\left(\hat{f}_{m,n}-\hat{f}_{m,n+1}\right)^2+\\ &\quad +w^s_{m,n}\left(\hat{f}_{m,n}-\hat{f}_{m+1,n}\right)^2+\\ &\quad +w^w_{m,n}\left(\hat{f}_{m,n}-\hat{f}_{m,n-1}\right)^2\Big]. \qquad (1)\end{aligned}$$

The weighting coefficients w can be seen as smoothing constraints, suppressing the variation in the specified direction (north, east, south and west). Using matrix notation, eq. (1) can be written as

$$\begin{aligned}\Psi(\hat{\mathbf{f}}) &= \frac{1}{2}\Big(\|\mathbf{S}_n\hat{\mathbf{f}}-\mathbf{b}_n\|^2+\|\mathbf{S}_e\hat{\mathbf{f}}-\mathbf{b}_e\|^2+\\ &\quad +\|\mathbf{S}_s\hat{\mathbf{f}}-\mathbf{b}_s\|^2+\|\mathbf{S}_w\hat{\mathbf{f}}-\mathbf{b}_w\|^2\Big). \qquad (2)\end{aligned}$$

Vector $\hat{\mathbf{f}}$ contains the estimated values $\hat{f}_{m,n}$ in arbitrary order. The matrix $\mathbf{S_{n,e,s,w}}$ is composed of the weighting coefficients $w_{n,e,s,w}$ according to the order of $\hat{\mathbf{f}}$. Vector $\mathbf{b}_{n,e,s,w}$ contains samples surrounding the lost block (*boundary pixels*) in the specific direction. The smoothness measure $\Psi(\hat{\mathbf{f}})$ gets minimal for $\hat{\mathbf{f}} = \hat{\mathbf{f}}_{\mathbf{opt}}$. Thus, the optimal estimation $\hat{\mathbf{f}}_{opt}$ can be found by

$$\frac{\partial\Psi}{\partial\hat{\mathbf{f}}_{opt}} = 0. \qquad (3)$$

3. HYBRID APPROACH

In this section we introduce our approach to the stereoscopic scenario. First we point out a method to retrieve the analogous block from the corresponding undamaged image and second we integrate the located block in a smoothing algorithm derived from the method described in section 2.

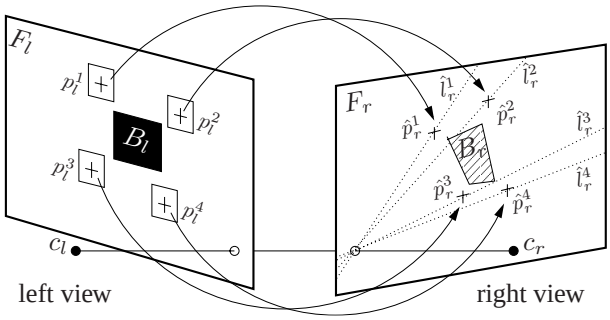

Fig. 2: Feature point correspondences for image warping

3.1. Block Search and Transformation

To determine the corresponding block we compute a 2D transformation to warp estimated pixels from the non erroneous image into the lost block.

The transformation of the erroneous region B_l in F_l is estimated with respect to the associated stereo image. This method is closely related to algorithms applied for *image mosaicing*. Using the projective transformation model [6], the image transformation $\mathcal{T}$ is represented by:

$$\begin{aligned}(x_r, y_r) &= \mathcal{T}(\mathbf{k};(x_l,y_l))\\ x_r &= \frac{a_1x_l+a_2y_l+a_3}{c_1x_l+c_2y_l+1} \qquad (4)\\ y_r &= \frac{b_1x_l+b_2y_l+b_3}{c_1x_l+c_2y_l+1}\end{aligned}$$

In these equations a position $\mathbf{p_l} = (x_l, y_l)$ in the erroneous frame is mapped to a position $\mathbf{p_r} = (x_r, y_r)$ in the undamaged frame. This transformation is controlled by a set of *warping parameters* $\mathbf{k} = [a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2]^T$.

The estimated parameter set $\hat{\mathbf{k}}$ is fully determined by four point correspondences between F_l and F_r. From F_l we extract *feature points* computed by the Harris corner detector [7], which is based on a gradient measurement. To find the correspondences a common feature block matching method is applied. As matching score the normalized cross correlation is used. We achieve half pixel accuracy.

Matching *feature points* is probably the most difficult problem. Therefore we use the advantages of epipolar geometry between both views, i.e. the reduction of the matching search area from the 2D image plane to a line, the so called epipolar line. For retrieving the epipolar lines, the fundamental matrix of the stereoscopic image pair, i.e. the algebraic representation of the epipolar geometry [8], is estimated. The projective mapping from points to lines is given by the following equation:

$$\mathbf{l_r^i} = \mathbf{F}\mathbf{p_l^i} \qquad i = 1\ldots4 \qquad (5)$$

$\mathbf{F}$ is the fundamental matrix and $\mathbf{l_r^i}$ is the epipolar line in the right image corresponding to the *feature point* $\mathbf{p_l^i}$ in the erroneous left image.

To reduce the probability of feature mismatching a computed feature correspondence $(\mathbf{p_l^i}, \mathbf{p_r^i})$ is only used if both appropriate points can be located on each others epipolar line and the matching score exceeds a minimum treshold. From the set of possible *feature points* the four $(\mathbf{p_l^1}, \ldots, \mathbf{p_l^4})$ next to the lost block are utilized. Fig. 2 illustrates the correspondences along the epipolar lines $(\mathbf{l_r^1}, \ldots, \mathbf{l_r^4})$.

Region B_r is warped with respect to the *warping parameters* $\hat{\mathbf{k}}$. The positions within the erroneous frame (x_l^j, y_l^j) are transformed into the positions (x_r^j, y_r^j) of the error free image F_r and pixel values are assigned.

$$F_l(x_l^j, y_l^j) = F_r(\mathcal{T}(\hat{\mathbf{k}}; (x_l^j, y_l^j))) \qquad (x_l^j, y_l^j) \in B_l \quad (6)$$

3.2. Block Smoothing

The corresponding block $\hat{B}_l$ found in section 3.1 is integrated in the smoothing measure $\Psi(\hat{\mathbf{f}})$, which was introduced in eq. (2). As already mentioned and illustrated in fig. 1, the adjacent pixels of $\hat{B}_l$ are treaded like additional boundary pixels of the lost block. The extended smoothing measure $\Psi_{ext}(\hat{\mathbf{f}})$ becomes

$$\begin{aligned}\Psi_{ext}(\hat{\mathbf{f}}) \;=\; & \frac{1}{2}\Big(\|\mathbf{S}_n\hat{\mathbf{f}} - \mathbf{b}_n\|^2 + \|\mathbf{S}_e\hat{\mathbf{f}} - \mathbf{b}_e\|^2 + \\ & + \|\mathbf{S}_s\hat{\mathbf{f}} - \mathbf{b}_s\|^2 + \|\mathbf{S}_w\hat{\mathbf{f}} - \mathbf{b}_w\|^2 + \\ & + \|\mathbf{S}_{stereo}\hat{\mathbf{f}} - \mathbf{b}_{stereo}\|^2\Big). \end{aligned} \quad (7)$$

Vector $\mathbf{b_{stereo}}$ is composed of the sample values of $\hat{B}_l$ arranged in the same order as $\hat{\mathbf{f}}$. Note that $\mathbf{b_{stereo}}$ contains N^2 nonzero values compared to N nonzero values of the boundary vectors.

Depending on the distance to the according border, we choose decreasing weighting coefficients within $\mathbf{S}_{n,e,s,w}$, while keeping a constant weight within $\mathbf{S}_{stereo}$. This adaption reduces discontinuities at the block border with less blurring.

The projective transformation, and thus the 3D block smoothing, would yield bad results if the four feature points have large variations in depth. Therefore the relative depth is estimated simply with triangulation. For depth variations larger than a given threshold, we apply only the monoscopic error concealment technique proposed in section 2.

4. SIMULATION RESULTS

To show the efficiency of our method, we take two example stereo image pairs (”Castle”, fig. 3 and ”Hall”, fig. 5). First, we assume several arbitrary block losses in the left images and reconstruct them with three different techniques: the monoscopic error concealment technique (section 2), the 2D warping with respect to the projective transformation model (section 3.1) and the 3D block smoothing described in section 3.2. For blocksizes of 8×8 and 16×16 the calculated block-PSNR is shown in table 1. The enhancement lies within 1dB and 3dB, although the real aim is to improve the stereoscopic depth perception.

Fig. 3: Erroneous left image (”Castle”)

Fig. 4: Left image, reconstructed with 3D-BS

	Hall		Castle	
PSNR in dB	size=8	size=16	size=8	size=16
Monoscopic	29.05	25.45	21.38	19.38
2D warping	29.50	28.52	22.79	20.21
3D-BS	30.18	28.43	23.06	20.87

Table 1. Average block-PSNR for arbitrary block losses

Fig. 3 shows the erroneous left image with 24 lost blocks of size 16 x 16. The 3D block smoothing is presented in fig. 4, which shows the reconstructed image. Due to the utilization of the second channel, the proposed algorithm yields a very high quality reconstruction result. Most notably the stereoscopic depth perception is much less affected than using monoscopic error concealment strategies. To strengthen these results, a psychovisual evaluation with

Fig. 5: Erroneous left image ("Hall")

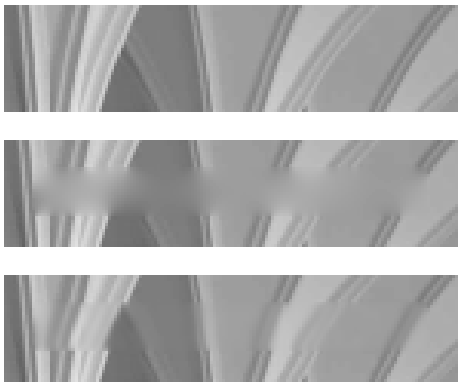

Fig. 6: Results of error concealment: original block (top), mono concealment (center), 3D-BS algorithm (bottom)

Fig. 7: Results of error concealment: original block (left), mono concealment (center), 3D-BS algorithm (right)

twelve test subjects was carried out.

Figures 5-7 show the concealment results of image pair "Hall" and illustrate the different features of the presented error concealment strategies. For a burst error of 8 consecutively lost 16×16 blocks the 3D-BS strategy yields better results than the monoscopic concealment strategy, which has principally low pass character. High frequency components of the lost block are reconstructed, but the 2D transformation might lead to deformation in case of discontinuities in depth. The block-PSNR values are 16.11dB (monoscopic) and 21.96dB (3D-BS) for the single lost block and 24.93dB (monoscopic) and 25.74dB (3D-BS) for the burst of $8 \times 16 \times 16$ blocks.

5. SUMMARY

In this paper we have presented a new approach for stereoscopic error concealment. Based on the projective transformation model and a monoscopic error concealment technique, we demonstrated the 3D block smoothing to improve reconstruction results of lost blocks in stereoscopic images. The stereoscopic depth perception is much less affected combining the good perceptual properties (*smoothness*) with additional information of the second frame than using only a monoscopic error concealment technique. The possibility of weighting the border and intersample variation differently makes the proposed algorithm very powerful. Even for small misalignments caused by modeling errors due to the warping process, a smooth border transition can be achieved. Further work include advanced methods to deal with discontinuities in depth and different estimation methods for the projective transformation coefficients.

6. REFERENCES

[1] G. Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, Feb 1992.

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

[3] W. IJsselsteijn, H. de Ridder, and J. Vliegen, "Subjective evaluation of stereoscopic images: effects of camera parameters and display duration," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. Vol.10, no. Iss.2, pp. 225–233, Mar 2000.

[4] L. Stelmach, W. Tam, D. Meegan, A. Vincent, and P. Corriveau, "Human perception of mismatched stereoscopic 3d inputs," in *International Conference on Image Processing*, vol. Vol. 1, 2000, pp. p. 5–8.

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

[6] 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.

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

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