TEMPORAL ERROR CONCEALMENT FOR FISHEYE VIDEO SEQUENCES BASED ON EQUISOLID RE-PROJECTION

Andrea Eichenseer, Jürgen Seiler, Michel Bätz, and André Kaup

Multimedia Communications and Signal Processing
Friedrich-Alexander University Erlangen-Nürnberg (FAU), Cauerstr. 7, 91058 Erlangen, Germany

ABSTRACT

Wide-angle video sequences obtained by fisheye cameras exhibit characteristics that may not very well comply with standard image and video processing techniques such as error concealment. This paper introduces a temporal error concealment technique designed for the inherent characteristics of equisolid fisheye video sequences by applying a re-projection into the equisolid domain after conducting part of the error concealment in the perspective domain. Combining this technique with conventional decoder motion vector estimation achieves average gains of 0.71 dB compared against pure decoder motion vector estimation for the test sequences used. Maximum gains amount to up to 2.04 dB for selected frames.

Index Terms— Error Concealment, Fisheye Lens, Temporal Prediction, Motion Vector Estimation

1. INTRODUCTION

Video surveillance, automotive, and also outdoor applications often make use of very wide fields of view (FOV) of 180 degrees and beyond. To capture such ultra wide-angle video sequences with a single camera, fisheye lenses [1] based on projection functions quite different from the pinhole model are employed. Many applications require the immediate coding of the obtained fisheye videos using a block-based hybrid video codec [2, 3], for instance. Subsequently transmitting the coded data from the camera to a receiver over error-prone channels may cause losses that the receiver side may want to conceal to reconstruct the visual quality to a certain degree.

Countless error concealment techniques can be found in literature, classified into three categories, namely spatial, temporal, and spatio-temporal techniques. Spatial error concealment techniques [4] rely only on information available in the video frame to be reconstructed. Temporal error concealment approaches like decoder motion vector estimation (DMVE) [5] exploit correlations within the temporal neighborhood of the distorted frame. The third category comprises spatio-temporal error concealment techniques which try to suitably combine spatial and temporal approaches [6]. Further improvement can be achieved by adding post-processing steps like denoising [7].

Fig. 1. Example frames of synthetically generated equisolid fisheye video sequences. Top: *Street*, bottom: *Room*.

In this paper, we consider temporal error concealment for fisheye video sequences. More specifically, we propose an adapted DMVE technique designed for equisolid fisheye data as depicted in Fig. 1. Block-matching error concealment techniques [8] are based on a translational motion model and are thus very much suited to rectilinear video data. For fisheye videos, however, the translational model is not a suitable assumption as they do not comply with the pinhole model. This was partly investigated in [9], where it was shown that inter-frame video coding and, consequently, traditional motion estimation works better in the perspective domain. As this observation can be extended towards block-matching error concealment methods like DMVE, our equisolid temporal error concealment (E-TEC) technique is based on a transform into the perspective domain to exploit its better suitability to the translational motion model. Following the motion search in the perspective domain, E-TEC employs a re-projection into the equisolid domain, where the actual concealment is conducted. E-TEC thus adapts the motion estimation technique described in [10] for use in temporal error concealment.
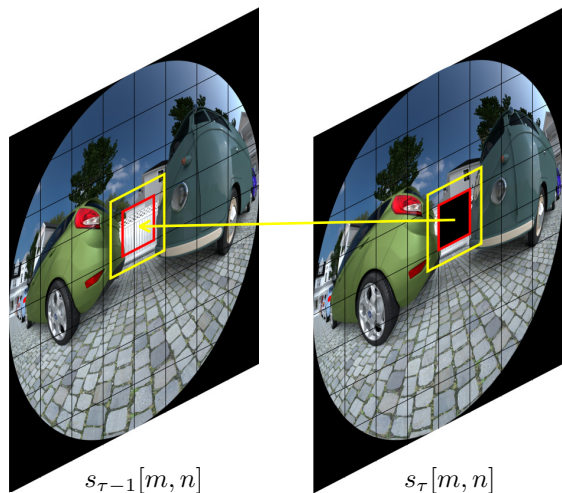
Fig. 2. DMVE example with decision area $\mathcal{D}$ and corresponding motion vector depicted in yellow and loss area $\mathcal{L}$ depicted in red.

2. DECODER MOTION VECTOR ESTIMATION

In the following, the principle of DMVE [5] is briefly outlined. Fig. 2 visualizes the approach. Given a video frame $s_\tau[m, n]$ at time $t = \tau$ containing a block loss, an error concealed frame $\tilde{s}_\tau[m, n]$ is obtained by:

$$\tilde{s}_\tau[m, n] = \begin{cases} s_{\tau-1}[m + \Delta m, n + \Delta n], & \forall (m, n) \in \mathcal{L} \\ s_\tau[m, n], & \forall (m, n) \notin \mathcal{L} . \end{cases} \quad (1)$$

Here, $\mathcal{L}$ describes the area of the lost block and $(\Delta m, \Delta n)$ denotes the motion vector used for concealing this block. The optimum motion vector is selected from a set of motion vector candidates $(\Delta m_i, \Delta n_i)$ defined within a certain search range and based on an error criterion such as the sum of squared differences (SSD):

$$\text{SSD}_i = \sum_{(m,n) \in \mathcal{D}} (s_\tau[m, n] - s_{\tau-1}[m + \Delta m_i, n + \Delta n_i])^2 . \quad (2)$$

$\mathcal{D}$ describes a decision area around the lost block, excluding the loss area $\mathcal{L}$. In Fig. 2, $\mathcal{D}$ comprises the area within the yellow block without the area of the red block. Minimizing SSD_i yields the motion vector to be used for concealing $\mathcal{L}$:

$$(\Delta m, \Delta n) = \underset{(\Delta m_i, \Delta n_i)}{\text{argmin}} \ \text{SSD}_i . \quad (3)$$

After obtaining the motion vector, the lost block can be substituted and thereby concealed by copying the corresponding block shifted by the motion vector from the reference frame $s_{\tau-1}[m, n]$ into the distorted frame as defined in (1).

Just like conventional block-based motion estimation [11] methods, DMVE relies on a translational motion model as this describes the predominant kind of motion in a typical video sequence. Since fisheye images are not based on a perspective projection function, they exhibit characteristics for which this model no longer holds true. We hence propose taking into account the projection function of fisheye images and introduce an adapted temporal error concealment method.
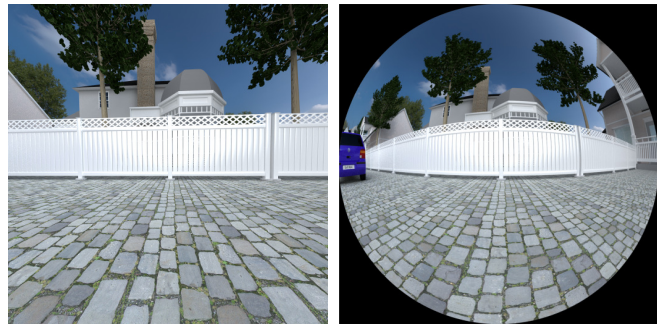

Fig. 3. Comparison of a perspective image (left) to its equisolid fisheye version (right) using the same spatial resolution.

3. TEMPORAL ERROR CONCEALMENT VIA EQUISOLID RE-PROJECTION

The different projection functions and resulting image characteristics of perspective images and equisolid fisheye images become quite evident by regarding Fig. 3. While the left image is obtained using the pinhole model

$$r_\text{p} = f \tan \theta , \quad (4)$$

i. e., perspective projection, the right one is based on equisolid projection [1]:

$$r_\text{e} = 2f \sin(\theta/2) . \quad (5)$$

In both cases, θ is the incident angle of light and f denotes the focal length. r_p and r_e describe the distance to the image center in the perspective and equisolid image, respectively. Using polar coordinates $(r_\text{p}, \phi_\text{p})$ and $(r_\text{e}, \phi_\text{e})$, r_p and r_e represent the radius, while ϕ_p and ϕ_e denote the angle. Evidently, equisolid projection allows a much larger FOV, but the resulting image no longer follows the rules of projective geometry and straight lines are mapped onto arcs. As a consequence, image processing techniques based on a translational motion model must be considered suboptimal as concluded in [9].

We hence propose an equisolid temporal error concealment (E-TEC) technique based on DMVE which conducts the motion vector search in the perspective domain [10]. Since projecting the entire equisolid image into the perspective domain is practically infeasible due to the vast amount of pixels this would result in, we instead manipulate the image coordinates $(r_\text{e}, \phi_\text{e})$ in a suitable fashion. We thus use

$$r_\text{p} = f \tan \left(2 \arcsin \left(\frac{r_\text{e}}{2f} \right) \right) \quad (6)$$

to back-project the image coordinates into the perspective domain $\mathcal{P}$. Since the translational model holds here, the addition of the motion vector candidate $(\Delta m_i, \Delta n_i)$ is conducted in this domain using a Cartesian representation. Afterwards, the now shifted polar coordinates $(r'_\text{p}, \phi'_\text{p})$ are re-projected into the equisolid domain $\mathcal{E}$ via

$$r'_\text{e} = 2f \sin \left(\frac{1}{2} \arctan \left(\frac{r'_\text{p}}{f} \right) \right) \quad (7)$$
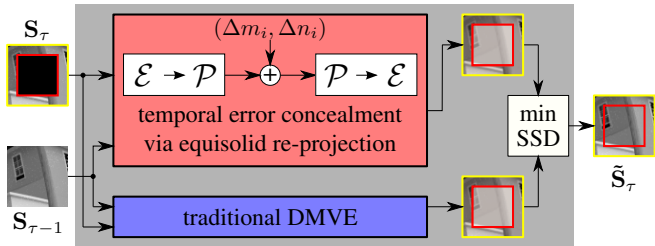
Fig. 4. Schematic depiction of HE-TEC (gray) combining the proposed E-TEC (light red) with conventional DMVE (blue).

and subsequently applied to a suitably upsampled and interpolated version of the reference frame to extract the corresponding pixel values. Note that in neither (6) nor (7) the angle is changed in any way, so that $\phi_{\mathrm{p}} = \phi_{\mathrm{e}}$ and $\phi_{\mathrm{e}}' = \phi_{\mathrm{p}}'$. The upsampling and interpolation of the reference frame is necessary since the Cartesian coordinates corresponding to $(r_{\mathrm{e}}', \phi_{\mathrm{e}}')$ are no longer comprised of integer values. To preserve a certain degree of accuracy, a suitable upsampling factor must thus be chosen, e. g., a factor of 8 for eighth-pixel accuracy.

Apart from the additional projections described above, the principle of the motion vector search, including the minimization of the SSD based on a decision area $\mathcal{D}$ around the lost block, is the same as for regular DMVE. To minimize the SSD of the decision area $\mathcal{D}$ and thus determine the motion vector $(\Delta m, \Delta n)$, all image coordinates $(r_{\mathrm{e}}, \phi_{\mathrm{e}}) \in \mathcal{D}$ are projected into the perspective domain, where the motion vector candidate $(\Delta m_i, \Delta n_i)$ is added. The resulting shifted coordinates are subsequently re-projected into the equisolid domain, where they can be applied to the reference frame, thus extracting the pixel values to be compared to $\mathcal{D}$. Repeating this for all motion vector candidates within the search range finally yields $(\Delta m, \Delta n)$. Having determined the motion vector $(\Delta m, \Delta n)$, the block to be used for concealing the loss area is obtained by projecting all image coordinates $(r_{\mathrm{e}}, \phi_{\mathrm{e}}) \in \mathcal{L}$ into the perspective domain, adding the motion vector, and re-projecting the shifted coordinates into the equisolid domain. Applying the image coordinates thus obtained to the upsampled reference frame then yields the block used for concealing the area $\mathcal{L}$ of the regarded lost block.

To implement the presented E-TEC method, we directly build upon conventional DMVE and thereby create a hybrid equisolid temporal error concealment (HE-TEC) technique. HE-TEC allows DMVE as an optional technique for the concealment of blocks where our E-TEC method meets its limits. One such limiting factor is the inverse tangent function in (7). As we define an integer-pixel search range in the perspective domain, employing (7) leads to a shortened search range in the equisolid domain. Since the original search range in the perspective domain is able to cover a larger range of motion, DMVE may outperform E-TEC in the case of fast motion or for lost blocks in the periphery of the fisheye image, as it is not inhibited by a shortened search range. This is especially true when nearing the 180 degree boundary of the fisheye image as these coordinates are located near infinity in the perspec-

Fig. 5. Example frames of real-world fisheye video sequences. Top left to bottom right: *Video1*, *Video2*, *Video3*, and *Video4*.

Sequence name	Resolution (pixels)	Frames per second	Length (frames)	Frames tested
Street	1088×1088	25	1800	55
Room	1088×1088	25	400	7
Video1	768×1216	30	30	29
Video2	768×1216	15	30	29
Video3	768×1216	30	30	29
Video4	768×1216	30	30	29

Table 1. General information on the test sequences.

tive domain. Any search range is consequently re-projected onto a very small area or even a single point in the equisolid domain according to the inverse tangent in (7), hence being unable to capture any kind of motion between frames.

For the proposed HE-TEC, we thus incorporate an SSD-based decision between pure DMVE and our equisolid re-projection variant E-TEC. HE-TEC is schematically depicted in Fig. 4, where $\mathbf{S}_\tau$, $\mathbf{S}_{\tau-1}$, and $\tilde{\mathbf{S}}_\tau$ denote the lossy signal to be concealed, the reference frame, and the error concealed signal, respectively. The blue box denotes the conventional DMVE approach. The light red box describes the proposed equisolid re-projection approach E-TEC. While only the projection of the image coordinates into the perspective domain $\mathcal{P}$, the motion vector addition, and the re-projection of the translated image coordinates into the equisolid domain $\mathcal{E}$ are explicitly visualized, any other necessary processing steps like the motion vector search based on SSD minimization, the upsampling of the reference frame, as well as the concealment block extraction are also part of E-TEC. Note that the projections do not require any information about actual pixel values as they work solely with image coordinates. In the following, HE-TEC is compared against pure DMVE.

4. SIMULATION SETUP AND RESULTS

To test our HE-TEC approach, we generated synthetic fisheye video sequences using blender [12] and, to that end, made use of several object models from [13] to create realistic scenes.
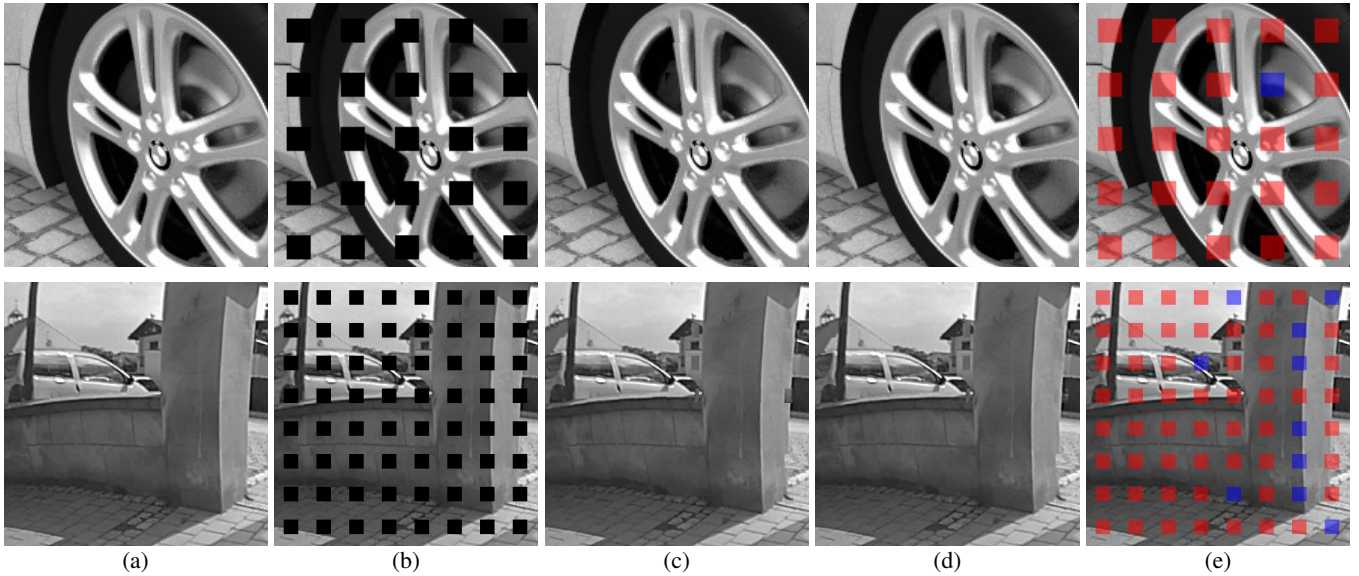
Fig. 6. Exemplary image detail results for *Street* (top) and *Video1* (bottom). (a) original image, (b) lossy image, (c) pure DMVE, (d) HE-TEC, (e) HE-TEC overlaid with red when the E-TEC approach was given preference and blue when conventional DMVE was performed.

	DMVE	HE-TEC	Gain	Gain$_{max}$
Street	29.62 dB	30.69 dB	**1.07 dB**	2.04 dB
Room	44.64 dB	45.06 dB	**0.42 dB**	0.78 dB
Video1	30.96 dB	31.62 dB	**0.66 dB**	1.48 dB
Video2	24.76 dB	25.47 dB	**0.71 dB**	1.03 dB
Video3	28.53 dB	29.18 dB	**0.65 dB**	0.87 dB
Video4	28.88 dB	27.65 dB	**0.77 dB**	1.67 dB
Average			0.71 dB	

Table 2. Average luminance PSNR results. In addition, the overall maximum gain achieved for a selected frame is given.

The blender setting for the camera was panoramic fisheye using equisolid projection. The FOV was set to 185 degrees, the focal length to 1.8 mm, and the sensor size to 5.2 mm by 5.2 mm, so that the entire circular fisheye can be captured. Fig. 1 shows exemplary frames of our synthetic fisheye sequences. *Street* employs a moving camera and static objects and the contained motion is mostly translational. *Room* on the other hand uses a static camera and various moving objects so that there is no global motion.

To further test HE-TEC on real-world video sequences, we used four traffic sequences which all contain global translational motion. For each real-world sequence, an example frame is depicted in Fig. 5. Regarding the FOV and focal length, we assume the same values as used for the synthetic sequences. Only the sensor size was changed to 4.6 mm by 2.9 mm, as the real-world sequences evidently consist of full-frame fisheye images which fill the entire sensor area. The sensor size was estimated by searching the maximum radius that was mapped onto the image plane along with the assumption that 5.2 mm is enough to represent the entire circular fisheye. Further information on the test sequences used is given in Table 1.

In all of the tests conducted, a fixed integer-pixel search range of 128 pixels in every direction was used for both regular DMVE as well as our HE-TEC technique so that 257×257 motion vector candidates were evaluated for each lost block. For HE-TEC, the re-projected image coordinates were applied to a reference frame upsampled by a factor of 8 using cubic convolution interpolation. The proposed HE-TEC technique was evaluated for multiple isolated block losses of 16×16 pixels throughout all tests conducted. The decision area $\mathcal{D}$ around each lost block was set to a width of 8 pixels so that the union $\mathcal{D} \cup \mathcal{L}$ forms an area of 32×32 pixels.

Table 2 summarizes the average luminance PSNR results calculated for the loss areas as well as the average gains obtained for each sequence. Additionally, the maximum gains achieved for selected frames of each sequence are given. For *Street*, average gains amount to 1.07 dB with an overall maximum of 2.04 dB. This result shows that equisolid re-projection is a suitable means for an improved motion search in fisheye sequences. Not surprisingly, there is a much lower gain for the static camera sequence *Room* since DMVE is able to achieve perfect signal reconstruction for most of the lost blocks and HE-TEC cannot improve on that. Nonetheless, small gains can be achieved if lost blocks contain parts of the few moving objects within this sequence.

In terms of real-world fisheye sequences, average gains of around 0.7 dB are obtained, showing that the proposed HE-TEC technique also works on non-synthetic sequences. Although the assumption of an equisolid projection is certainly not an accurate one, it is good enough to achieve an improved concealment result. Adapting HE-TEC to calibrated projection information should further increase the gain. As mentioned earlier, the search range is still a limiting factor of HE-TEC, so that an adaptation considering the equisolid projection should also potentially increase the obtained gains.

A zoomed-in visual example is given in Fig. 6. (b) shows the error pattern, (c) the concealment results obtained by DMVE, and (d) the HE-TEC results. In (e), the HE-TEC results are additionally overlaid with a decision mask. Red color denotes those lost blocks for which error concealment was done using the projection-based E-TEC approach, while blue color denotes blocks for which conventional DMVE was chosen. Although the overall visual impression seems very similar for both DMVE and HE-TEC, differences along curved shapes can be made out upon closer inspection. Here, the equisolid re-projection technique is able to achieve better reconstruction results that sum up to an improved image quality. These visual results are representative for all frames tested, synthetic and real-world data alike.

When evaluating the HE-TEC results for the real-world sequences, it was observed that on average, 75 % of all lost blocks were concealed via E-TEC, i.e., via equisolid re-projection, while conventional DMVE was chosen for only 25 % of the blocks. For *Street* and *Room*, E-TEC was employed for 71 % and 96 % of all lost blocks, respectively. It is quite evident that our equisolid re-projection technique is chosen for most of the lost blocks, thus substantiating the PSNR results. Since the majority of lost blocks is concealed by the introduced E-TEC approach, its implementation as a stand-alone error concealment method is also conceivable, rendering the blue box as well as the SSD-based decision in Fig. 4 obsolete.

5. CONCLUSION

In this paper, we introduced a temporal error concealment technique for fisheye video sequences via equisolid re-projection. Based on the knowledge that the translational motion model does not hold for fisheye videos due to the different underlying projection function, we employed suitable projections from the equisolid to the perspective domain and vice versa in order to conceal the lost blocks with the help of a reference frame. Furthermore, a hybrid technique combining this approach with conventional DMVE was proposed and evaluated. Average gains in luminance PSNR amounted to 0.71 dB for both the synthetic and real-world sequences tested, letting us conclude that exploiting knowledge about optics that differ from the conventional pinhole model is a suitable means for improving image reconstruction quality.

The development of the proposed E-TEC method as a stand-alone technique is part of work in progress. A major point of interest to that end is the suitable handling of the peripheral parts of circular fisheye frames, i.e., those parts, where the FOV gets close to and surpasses the 180 degree boundary. Current work also investigates spatio-temporal error concealment for fisheye video sequences as well as optimizations with regard to the motion search. Of particular interest here is the reduction of motion vector candidates to evaluate.

REFERENCES

[1] K. Miyamoto, "Fish Eye Lens," *Journal of the Optical Society of America*, vol. 54, no. 8, pp. 1060–1061, August 1964.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[3] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, December 2012.

[4] X. Li, "Variational Bayesian Image Processing on Stochastic Factor Graphs," in *Proc. IEEE Int. Conf. on Image Processing*, San Diego, USA, October 2008, pp. 1748–1751.

[5] J. Zhang, J. F. Arnold, and M. R. Frater, "A Cell-Loss Concealment Technique for MPEG-2 Coded Video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no. 4, pp. 659–665, June 2000.

[6] M.-C. Hwang, J.-H. Kim, C.-S. Park, and S.-J. Ko, "Improved Fading Scheme for Spatio-Temporal Error Concealment in Video Transmission," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E91-A, no. 3, pp. 740–748, March 2008.

[7] J. Seiler, M. Schöberl, and A. Kaup, "Spatio-Temporal Error Concealment in Video by Denoised Temporal Extrapolation Refinement," in *Proc. IEEE Int. Conf. Image Processing*, Melbourne, Australia, September 2013, pp. 1613–1616.

[8] S. Tsekeridou and I. Pitas, "MPEG-2 Error Concealment Based on Block-Matching Principles," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no. 4, pp. 646–658, June 2000.

[9] A. Eichenseer and A. Kaup, "Coding of Distortion-Corrected Fisheye Video Sequences Using H.265/HEVC," in *Proc. IEEE Int. Conf. on Image Processing*, Paris, France, October 2014, pp. 4132–4136.

[10] A. Eichenseer, M. Bätz, J. Seiler, and A. Kaup, "A Hybrid Motion Estimation Technique for Fisheye Video Sequences Based on Equisolid Re-Projection," to appear in *Proc. IEEE Int. Conf. on Image Processing*, Quebec City, Canada, September 2015.

[11] M. Santamaria and M. Trujillo, "A Comparison of Block-Matching Motion Estimation Algorithms," in *Proc. IEEE 7th Colombian Computing Congress*, Medellin, Colombia, October 2012, pp. 1–6.

[12] "Blender Version 2.71," http://www.blender.org.

[13] "Blend Swap," http://www.blendswap.com.