A 3D-TV Approach Using Depth-Image-Based Rendering (DIBR)

CHRISTOPH FEHN
Image Processing Department
Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz Institut
Einsteinufer 37, 10587 Berlin, Germany
tel.: +49 – (0)30 31002611, fax: +49 – (0)30 3927200
email: christoph.fehn@hhi.fraunhofer.de

ABSTRACT
This paper will present details of a system that allows for an evolutionary introduction of depth perception into the existing 2D digital TV framework. The work is part of the European Information Society Technologies (IST) project "Advanced Three-Dimensional Television System Technologies" (ATTEST), an activity, where industries, research centers and universities have joined forces to design a backwards-compatible, flexible and modular broadcast 3D-TV system [1]. In contrast to former proposals, which often relied on the basic concept of "stereoscopic" video, this new idea is based on a more flexible joint transmission of monoscopic video and associated per-pixel depth information. From this data representation, one or more "virtual" views of the 3D scene can then be synthesized in real-time at the receiver side by means of so-called depth-image-based rendering (DIBR) techniques. This paper (a) highlights the advantages of this new approach on 3D-TV and (b) develops an efficient algorithm for the generation of "virtual" 3D views that can be reproduced on any stereoscopic- or autostereoscopic 3D-TV display.

KEY WORDS
Image Processing and Analysis, Depth-Image-Based Rendering (DIBR), 3D Image Warping, Stereoscopy, 3D-TV

1 A New Approach on 3D-TV

The ambitious aim of the European IST project ATTEST is to design a novel, backwards-compatible and flexible broadcast 3D-TV system [1]. In contrast to former proposals, which often relied on the basic concept of "stereoscopic" video, i. e. the capturing, transmission and display of two separate video streams – one for the left eye and one for the right eye –, this new idea is based on a more flexible joint transmission of monoscopic video and associated per-pixel depth information (see Fig. 1). From this data representation, one or more "virtual" views of the 3D scene can then be generated in real-time at the receiver side by means of so-called depth-image-based rendering (DIBR) techniques.

Compared with the classical approach of "stereoscopic" video, the proposed 3D-TV system has a number of advantages that can be shortly summarized as follows:

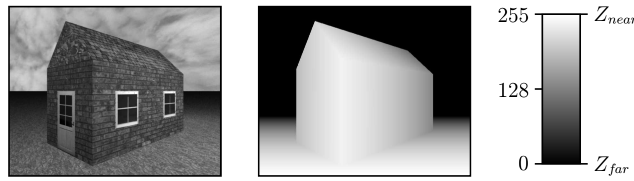


Figure 1. **The ATTEST data representation format.** It consists of: (a) Regular 2D color video; (b) Associated 8-bit depth-images that are normalized to a near clipping plane Z_{near} and a far clipping plane Z_{far}.

+ The 3D reproduction can be adjusted to a wide range of different stereoscopic displays and projection systems. As the required left- and right-eye views are only generated at the 3D-TV receiver, their appearance in terms of 'perceived depth' can be adapted to the particular viewing conditions. This allows to provide the viewer with a customized 3D experience that is comfortable to watch on any kind of stereoscopic- or autostereoscopic 3D-TV display [2].

+ 2D-to-3D conversion techniques based on "structure from motion" approaches can be used to generate the required depth information for already recorded monoscopic video material [3, 4, 5]. This is a very important point, as it seems clear that the success of any future 3D-TV broadcast system will depend to a great extent on the timely availability of sufficient interesting and exciting 3D video material.

+ Head-motion parallax (HMP) can be supported to provide an additional extrastereoscopic depth cue (see Section 3.3). This also eliminates the well-known "shear-distortions" that are usually experienced with stereoscopic- or autostereoscopic 3D-TV systems [6]. In addition to that, this feature can also be used to provide an increased sensation of depth on conventional, monoscopic 2D-TV displays [2].

+ The system allows the viewer to adjust the reproduction of depth to suit his/her personal preferences (see Section 3.2) – much like every conventional 2D-TV set allows the viewer to adjust the color reproduction by means of a (de-)saturization control. This is an important system feature taken into account the fact that there is a difference in depth appreciation over age groups. A recent study by Norman *et al.* for example demonstrated that older adults were less sensitive than younger adults to perceiving stereoscopic depth, in particular when screen parallax (see Section 3) was higher [8].

+ Photometrical asymmetries, e. g. in terms of brightness, contrast or color, between the left- and the right-eye view, which can destroy the stereoscopic sensation [7], are eliminated from the first, as both views are effectively synthesized from the same original image.

+ The ATTEST data representation format of monoscopic video plus associated per-pixel depth information is ideally suited to facilitate 3D post-processing. It enables automatic object segmentation based on depth-keying and allows for an easy integration of synthetic 3D objects into "real-world" sequences (augmented reality) [9]. This is an important prerequisite for advanced television features such as 'virtual advertisement' as well as for all kinds of real-time 3D special effects.

+ The provided per-pixel depth information also allows – in principle – to simulate the *depth of field* effect of natural vision during the synthesis of the "virtual" stereoscopic images. According to preliminary experiments conducted by Blohm *et al.*, such an advanced feature would be well suited to additionally improve the 3D viewing comfort on so-called "depth-of-interest" displays [10].

The remainder of this paper concentrates on the development of an efficient depth-image-based rendering (DIBR) algorithm that can be used for the real-time generation of "virtual" stereoscopic 3D images.

2 Depth-Image Based Rendering (DIBR)

Depth-image-based rendering (DIBR) is the process of synthesizing "virtual" views of a scene from still or moving images and associated per-pixel depth information [11, 12]. Conceptually, this novel view generation can be understood as the following two-step process: At first, the original image points are reprojected into the 3D world, utilizing the respective depth data. Thereafter, these 3D space points are projected into the image plane of a "virtual" camera, which is located at the required viewing position. The concatenation of reprojection (2D-to-3D) and subsequent projection (3D-to-2D) is usually called 3D image warping in the Computer Graphics (CG) literature and will be derived very briefly in the following.

2.1 3D Image Warping

Consider a system of two cameras and an arbitrary 3D space point M with the projections m and m' in the first-, resp. the second view. Under the assumption that the world coordinate system equals the camera coordinate system of the first camera[1], the two *perspective projection* equations result to:

$$\tilde{\mathbf{m}} \cong \mathbf{A}\mathbf{P}_n\tilde{\mathbf{M}} \quad (1)$$

$$\tilde{\mathbf{m}}' \cong \mathbf{A}'\mathbf{P}_n\mathbf{D}\tilde{\mathbf{M}} \ , \quad (2)$$

where $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{m}}'$, resp. $\tilde{\mathbf{M}}$ symbolize the two 2D image points, resp. the 3D space point in homogeneous notion and the symbol $\cong$ denotes 'equality up to a non-zero scale-factor' [4]. The 4×4 matrix $\mathbf{D}$ contains the rotation $\mathbf{R}$ and the translation $\mathbf{t}$ that transform the 3D point from the world coordinate system into the camera coordinate system of the second view and the 3×3 matrices $\mathbf{A}$ and $\mathbf{A}'$ specify the intrinsic parameters of the first-, resp. the second camera. Finally, the 3×4 identity matrix $\mathbf{P}_n$ designates the so-called normalized perspective projection matrix [13].

Rearranging (1) gives an affine representation of the 3D space point M that is, however, still dependent on its depth value Z:

$$\mathbf{M} = Z\mathbf{A}^{-1}\tilde{\mathbf{m}} \ . \quad (3)$$

Substituting (3) into (2) then leads to the classical affine *disparity equation*, which defines the depth-dependent relation between corresponding points in two perspective images of the same 3D scene:

$$Z'\tilde{\mathbf{m}}' = Z\mathbf{A}'\mathbf{R}\mathbf{A}^{-1}\tilde{\mathbf{m}} + \mathbf{A}'\mathbf{t} \ . \quad (4)$$

This disparity equation can also be considered as a *3D image warping* formalism, which can be used to generate an arbitrary novel view from a known reference image. This only requires the definition of the position and orientation of a "virtual" camera relative to the reference camera as well as the declaration of the "virtual" camera's intrinsic parameters. Then, if the depth values of the corresponding 3D space points are known for every pixel of the original image, the "virtual" view can be synthesized by applying (4) to all original image points. (The "real-world" problems of image resampling, resolving the visibility problem and handling of disocclusions in the novel view will be considered later in Section 4.)

3 Stereoscopic Image Creation

On a stereoscopic- or autostereoscopic 3D-TV display, two slightly different perspective views of a 3D scene are reproduced (quasi-)simultaneously on a joint image plane (see Fig. 2). The horizontal differences between these left- and right-eye views, the so-called *screen parallax* values, are interpreted by the human brain and the two images are fused into a single, three-dimensional percept.

[1] This special choice of the world coordinate system doesn't limit the universality of the following expressions, it just simplifies the mathematical formalism.

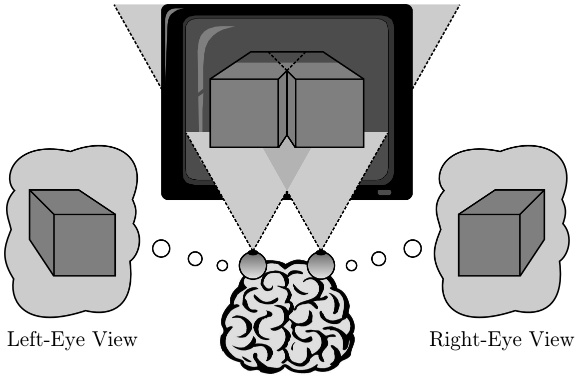


Figure 2. **Binocular depth reproduction on a stereoscopic 3D-TV display.** Two different perspective views, i. e. one for the left eye and one for the right eye, are reproduced (quasi-)simultaneously on a joint image plane.

In the ATTEST approach on 3D-TV, such stereoscopic images are not captured directly with a "real" stereo camera, rather they are synthesized from monoscopic video and associated per-pixel depth information. How this is done in a simple and effective way will be described in detail in the following.

3.1 Shift-Sensor Algorithm

In "real", high-quality stereo cameras, usually one of two different methods is utilized to establish the so-called *zero-parallax setting (ZPS)*, i. e. to choose the *convergence distance* Z_c in the 3D scene [14]. In the "toed-in" approach, the ZPS is chosen by a joint inward-rotation of the left- and right-eye camera. In the so-called *shift-sensor* approach, a plane of convergence is established by a small shift h of the parallel positioned camera's CCD sensors (see Fig. 3).

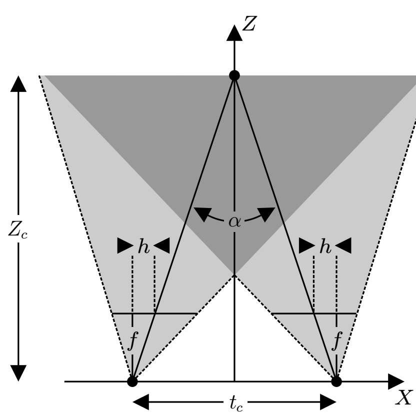


Figure 3. **Shift-sensor stereo camera setup.** In a shift-sensor stereo camera setup, the convergence distance Z_c is established by a shift h of the camera's CCD sensors.

While, technically, the "toed-in" approach is easier to realize in "real" stereo cameras, the shift-sensor approach is usually prefered because it doesn't introduce unwanted vertical differences – which are known to be a potential source of eye-strain – between the left- and the right-eye view [15]. Fortunately, this method is actually easier to implement with depth-image-based rendering (DIBR) as the required signal processing is only one-dimensional. All that is needed is the definition of two "virtual" cameras – one for the left eye and one for the right eye. With respect to the original view, these cameras are symmetrically displaced and their CCD sensors are shifted relative to the position of the lenses. Mathematically, this sensor shift can be formulated as a displacement of a camera's principal point c [13]. The intrinsic parameters of the two "virtual" cameras are therefore chosen to exactly correspond to the intrinsic camera parameters of the original view except for the horizontal shift h of the respective principal point. This can also be written as:

$$\mathbf{A}^* = \mathbf{A} + \begin{bmatrix} 0 & 0 & h \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} , \tag{5}$$

where the asterix symbol, which is used as a superscript here and in the following, should be substituted by either a single- or a double dash, e. g. $\mathbf{A}^*$ means either $\mathbf{A}'$ or $\mathbf{A}''$, to denote that the equation specifies the intrinsic parameters of either the left- or the right "virtual" camera.

Using the expression in (5) and taking into account that the movement of the two "virtual" cameras is restricted to be only translational with respect to the reference camera, i. e. $\mathbf{R} = \mathbf{I}$, where $\mathbf{I}$ is the 3×3 identity matrix, the following simplifications can be made in the general 3D warping equation (4):

$$\mathbf{A}^*\mathbf{R}\mathbf{A}^{-1} = \mathbf{A}^*\mathbf{A}^{-1} = \mathbf{I} + \begin{bmatrix} 0 & 0 & h \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} . \tag{6}$$

Inserting the simplified expression (6) into (4) yields the following reduced form of the general 3D warping equation:

$$Z^*\tilde{\mathbf{m}}^* = Z \left(\tilde{\mathbf{m}} + \begin{bmatrix} h \\ 0 \\ 0 \end{bmatrix} \right) + \mathbf{A}^*\mathbf{t} . \tag{7}$$

This expression can be simplified even more by taking into account that the only non-zero translational component needed to create a "virtual" shift-sensor camera setup is a horizontal translation t_x inside the focal plane of the original camera. With $t_z = 0$, it follows that the depth value of a 3D space point is the same in the world coordinate system – which was chosen to equal the camera coordinate system of the original view – and in the coordinate system of the "virtual" camera, i. e. $Z^* = Z$. Therefore (7) further reduces to:

$$\tilde{\mathbf{m}}^* = \tilde{\mathbf{m}} + \frac{\mathbf{A}^*\mathbf{t}}{Z} + \begin{bmatrix} h \\ 0 \\ 0 \end{bmatrix} \quad \text{with } \mathbf{t} = \begin{bmatrix} t_x \\ 0 \\ 0 \end{bmatrix} . \tag{8}$$

In this case, the affine pixel position (u, v) of each warped image point can simply be calculated as:

$$u^* = u + \frac{\alpha_u t_x}{Z} + h \quad \text{, resp.} \quad v^* = v \ . \tag{9}$$

The horizontal camera translation t_x is defined to equal the half of the chosen *interaxial distance* t_c[2], with the direction of the movement given by:

$$t_x = \begin{cases} -\frac{t_c}{2} & : \quad \text{left-eye view} \\ +\frac{t_c}{2} & : \quad \text{right-eye view} \ . \end{cases} \tag{10}$$

As already described, the amount of the sensor shift h depends on the selected convergence distance Z_c and can be calculated by taking into account that for $Z = Z_c$ the horizontal component u^* of the simplified 3D warping equation (9) must be the same in the left- and in the right view, i. e. $u' = u''$, which leads to the following simple expression:

$$h = -t_x \frac{\alpha_u}{Z_c} \ , \tag{11}$$

where t_x is also defined by (10).

The expressions in (9) to (11) fully define a simplified 3D warping equation that can be used to efficiently implement a “virtual” shift-sensor stereo camera setup. Table 1 shows, how the resulting 3D reproduction is influenced by the choice of the three main system variables, i. e. by the choice of the interaxial distance t_c, the focal length f of the reference camera and the convergence distance Z_c. The respective changes in screen parallax values, perceived depth and object size are qualitatively equal to what happens in a “real” stereo camera when the system parameters are manually adjusted.

Parameter	+/–	Scr. par.	Perc. depth	Obj. size
t_c	+	Increase	Increase	Constant
	–	Decrease	Decrease	Constant
f	+	Increase	Increase	Increase
	–	Decrease	Decrease	Decrease
Z_c	+	Decrease	Shift (fore)	Constant
	–	Increase	Shift (aft)	Constant

Table 1. **Effects of “virtual” stereo camera setup parameters.** Qualitative changes in screen parallax values, perceived depth and object size when varying the interaxial distance t_c, the focal length f or the convergence distance Z_c of the “virtual” stereo camera setup (after [16]).

3.2 Viewer-Control Over Depth Impression

It is a fact well-known among visual perception researchers that there is a difference in depth appreciation over age groups. A study by Norman *et al.* for example demonstrated that older adults were less sensitive than younger adults to perceiving stereoscopic depth, in particular when screen parallax was higher [8]. Furthermore, the ability to fuse stereoscopic images also seems to depend to some extend on the experience of the viewer with the three-dimensional medium. Experienced 3D-TV viewers seem to be able to fuse images with larger values of screen parallax than inexperienced viewers usually are [17]. Finally, it also seems to be a matter of personal taste whether or not a 3D video sequence is experienced as more enjoyfull when it provides very strong stereoscopic depth cues – and taste, as everyone knows, can differ greatly from one individual to the next. It is therefore an important feature of the described new approach on 3D-TV that the system is able to provide the viewer with the ability to adjust the reproduction of depth to suit his/her own personal preferences. In principle, an implementation could allow the viewer to individually readjust the two main stereoscopic parameters, i. e. the interaxial distance t_c as well as the convergence distance Z_c of the above-developed “virtual” shift-sensor stereo camera setup. This would give the viewer the possibility to increase or decrease the overall perceived depth as well as to shift the displayed 3D scene foreward and backwards in space (see again Table 1). However, changing the predefined convergence distance Z_c might result in a conflict between the stereoscopic depth cue and the monoscopic cue of interposition, which is experienced as very annoying by most people [7]. It therefore seems to be more appropriate – at least for fast moving images – to only provide the viewer with a means to adjust the interaxial distance t_c to change the overall perceived depth. Mathematically, this can easily be realized by substituting (10) with:

$$t_x = s \cdot \begin{cases} -\frac{t_c}{2} & : \quad \text{left-eye view} \\ +\frac{t_c}{2} & : \quad \text{right-eye view} \ , \end{cases} \tag{12}$$

where s is a scale factor that defaults to a value of $s = 1$ and that can be adjusted by the viewer between a minimum value $s_{min} = 0$ (resulting in no perceived depth at all) and an appropriate maximum value $s_{max} > 1$ (resulting in increased perceived depth).

3.3 Head-Motion Parallax

An important feature of the proposed new approach on 3D-TV is the ability to support head-motion parallax (HMP), i. e. to adjust the perspective of the synthesized “virtual” views to the actual viewing position of the user. This feature can be used to further enhance the perceived realism of stereoscopically displayed images by providing an additional, extrastereoscopic depth cue. This also eliminates the well-known “shear-distortions” that are usually experienced with stereoscopic 3D-TV systems [6]. In addition to that, head-motion parallax can also be used to provide an increased sensation of depth on conventional, monoscopic 2D-TV displays [2].

[2] This value is usually chosen to equal the average human eye separation of approximately 64 mm. However, for some 3D scenes a smaller or even larger t_c might be required to achieve the desired artistical 3D effect.

Supporting horizontal head-motion parallax in the shift-sensor algorithm described in Section 3.1 only requires to update the horizontal component u^* of the simplified 3D warping equation (9) with an additional translational term t_{hmp}:

$$u^* = u + \frac{\alpha_u \left(t_x + t_{hmp}\right)}{Z} + h \, , \qquad (13)$$

where t_x and h are the same as in (10) and (11).

4 Implementation Details

To verify the concept of the above-descibed "virtual" shift-sensor stereo camera setup, the algorithm was implemented in a real-time 3D-TV Player[3]. The following sections provide some implementation details that are specific to the current version of our software.

4.1 Visibility

During the generation of a "virtual" view, it can happen that two different original image points are warped to the same location in the new image. This situation occurs when one of the corresponding 3D space points is occluded by the other one in the novel view. A very simple way to resolve this *visibility problem* during the 3D warp is to process the pixel of the original image in a so-called *occlusion-compatible warp order* [12]. This processing order only depends on the relative positioning of the "virtual" camera with respect to the original camera and is therefore independent of the 3D scene itself. The effect of adhering to it is that closer points are always warped later, thus automatically overwriting points further away.

While in principle 18 different warp orders result for the most general form of the 3D warping equation (4), only two different processing orders have to be implementaed for the described "virtual" shift-sensor algorithm with horizontal head-motion parallax support. For the left-eye view, the columns of the original image are processed away from the right- towards the left image border. This warp order must simply be reversed for the right-eye view.

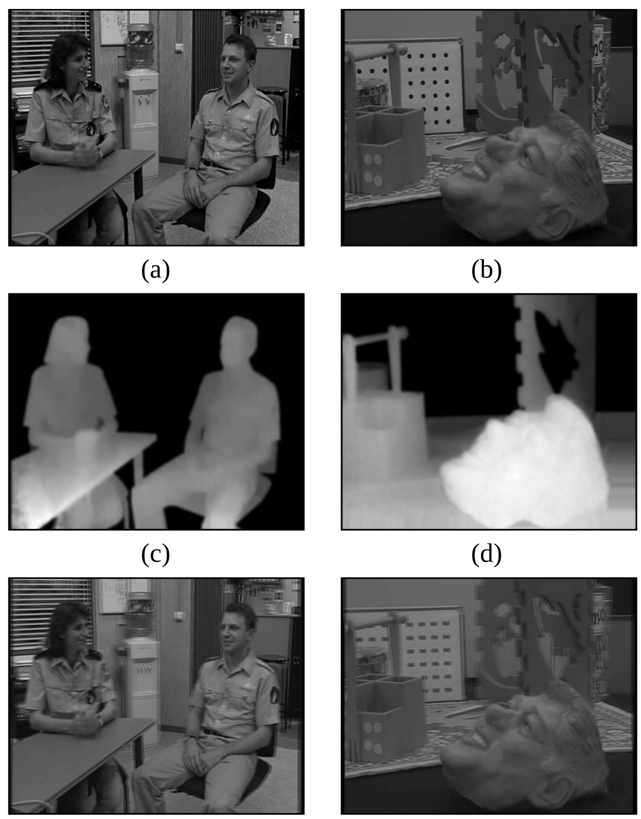
4.2 Disocclusions and Hole-Filling

One major problem of the described depth-image-based rendering (DIBR) algorithm is due to the fact that areas, which are occluded in the original view, might become visible in any of the "virtual" left- and right-eye views, an event that is usually referred to as *disocclusion* in the Computer Graphics (CG) literature. The questions results, how these disocclusions should be treated during the view synthesis, as information about the previously occluded areas is neither available in the monoscopic video nor in the accompanying depth-images?

Often described solutions to this problem try to 'fill-in' the disoccluded areas using either a simple foreground/background color interpolation or a somewhat more complex background color extrapolation, resp. mirroring technique. However, depending on the background texture, all these approaches are known to produce more or less annoying visible artifacts in the synthesized views. In ATTEST, a different way to handle disocclusions was found to provide superior synthesis results. The idea is to pre-process the depth values such that no disocclusions appear in the "virtual" stereoscopic views. Currently, we are only using a very simple 2D gaussian smoothing. Future work will concentrate on filtering algorithms that are only applied in areas of large horizontal depth discontinuities.

5 Experimental Results

Fig. 4 displays some experimental results. Each of the upper two images (a-b) shows a monoscopic video frame from the interlaced, TV resolution test sequences 'Interview' and 'Orbi'. The accompanying smoothed depth information can be found in the middle (c-d). The lower two images (e-f) show overlayed "virtual" left- and right-eye views that were generated in real-time on a 2.2 GHz PC.



(a) (b) (c) (d) (e) (f)

Figure 4. **Experimental results for the 'Interview' and 'Orbi' test sequences.** (a-b) Monoscopic video; (c-d) Accompanying smoothed depth information; (e-f) Overlayed "virtual" left- and right-eye views.

[3] A demo version of our 3D-TV Player software is available for free from the author of this publication.

The figure shows that the described shift-sensor algorithm is able to efficiently synthesize high-quality stereoscopic views of typical "real-world" scenes. With the "virtual" stereo camera setup parameters carefully chosen, the screen parallax values don't exceed a maximum of 3% of the image width. This limit is generally considered to be well suited to provide a visually pleasing 3D impression on stereoscopic- or autostereoscopic 3D-TV displays [18].

6 Conclusion

This paper described a new approach on 3D-TV using depth-image-based rendering (DIBR). A number of advantages over the classical approach of "stereoscopic" video, such as the possibility to support viewer-control over depth impression and head-motion parallax, have been pointed up and an actual real-time implementation based on a simple "virtual" shift-sensor algorithm has been provided. Finally, some experimental results have been presented, which prove the validity of the concept.

Acknowledgements

This work has been sponsored by the European Commission (EC) through their Information Society Technologies (IST) program under proposal No. IST-2001-34396. The author would like to thank the project officers as well as all project partners for their support.

References

[1] C. Fehn, P. Kauff, M. Op de Beeck, M. Ernst, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, and I. Sexton, An Evolutionary and Optimized Approach on 3D-TV, *Proceedings of International Broadcast Conference '02*, Amsterdam, The Netherlands, 2002, 357-365.

[2] C. Fehn and P. Kauff, Interactive Virtual View Video (IVVV) – The Bridge Between 3D-TV and Immersive TV, *Proceedings of SPIE Three-Dimensional TV, Video and Display '02*, Boston, MA, USA, 2002, 14-25.

[3] O. Faugeras, Q.-T. Luong, and T. Papadopoulo, *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*, MIT Press, Cambridge, MA, USA, 2001.

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

[5] M. Pollefeys, 3D Modelling from Images, Tutorial given at European Conference on Computer Vision '00, Dublin, Ireland, 2000.

[6] D. Runde, How to Realize a Natural Image Reproduction Using Stereoscopic Displays With Motion Parallax, *IEEE Transactions on Circuits and Systems for Video Technology*, 10(3), 2000, 376-386.

[7] L. Lipton, *Foundations of the Stereoscopic Cinema – A Study in Depth*, Van Nostrand Rheinhold, New York, NY, USA, 1982.

[8] J. Norman, T. Dawson, and A. Butler, The Effects of Age Upon the Perception of Depth and 3-D Shape From Differential Motion and Binocular Disparity, *Perception*, 29(11), 2000, 1335-1359.

[9] R. Gvili, A. Kaplan, E. Ofek, and G. Yahav, Depth Keying, *Proceedings of SPIE Electronic Imaging '03*, Santa Clara, CA, USA, 2003.

[10] W. Blohm, I. P. Beldie, K. Schenke, K. Fazel and S. Pastoor, Stereoscopic Image Representation with Synthetic Depth of Field, *Journal of the SID*, 33(5), 1997, 7-31.

[11] W. R. Mark, *Post-Rendering 3D Image Warping: Visibility, Reconstruction and Performance for Depth-Image Warping*, PhD thesis, University of North Carolina at Chapel Hill, 1999.

[12] L. McMillan, *An Image-Based Approach on Three-Dimensional Computer Graphics*, PhD thesis, University of North Carolina at Chapel Hill, 1997.

[13] G. Xu and Z. Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition*, Kluver Academic Publishers, Dordrecht, The Netherlands, 1996.

[14] A. Woods, T. Docherty, and R. Koch, Image Distortions in Stereoscopic Video Systems, *Proceedings of SPIE Stereoscopic Displays and Applications '93*, San Jose, CA, USA, 1993.

[15] W. A. IJsselsteijn, H. de Ridder, and J. Vliegen, Effects of Stereoscopic Filming Parameters and Display Duration on the Subjective Assessment of Eye Strain, *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems '00*, San Jose, CA, USA, 2000, 12-22.

[16] P. Milgram and M. Krüger, Adaptation Effects in Stereo Due to On-line Changes in Camera Configuration, *Proceedings of SPIE Stereoscopic Displays and Applications '92*, San Jose, CA, USA, 1992, 122-134.

[17] L. Lipton, Stereographics Developers' Handbook, 1997.

[18] C. W. Smith and A. A. Dumbreck, 3-D TV: The Practical Requirements, *Television: Journal of the Royal Television Society*, 1988, 9-15.