



New View Synthesis for Stereo Cinema by Hybrid Disparity Remapping

Frédéric Devernay, Sylvain Duchêne

► To cite this version:

Frédéric Devernay, Sylvain Duchêne. New View Synthesis for Stereo Cinema by Hybrid Disparity Remapping. International Conference on Image Processing (ICIP), Sep 2010, Hong Kong, Hong Kong SAR China. 10.1109/ICIP.2010.5649194 . inria-00531369

HAL Id: inria-00531369

<https://inria.hal.science/inria-00531369>

Submitted on 2 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NEW VIEW SYNTHESIS FOR STEREO CINEMA BY HYBRID DISPARITY REMAPPING

Frédéric Devernay, Sylvain Duchêne

INRIA Grenoble - Rhône-Alpes, France

ABSTRACT

The 3-D shape perceived from viewing a stereoscopic movie depends on the viewing conditions, most notably on the screen size and distance, and depth and size distortions appear because of the differences between the shooting and viewing geometries. When the shooting geometry is constrained, or when the same stereoscopic movie must be displayed with different viewing geometries (e.g. in a movie theater and on a 3DTV), these depth distortions may be reduced by new view synthesis techniques. They usually involve three steps: computing the stereo disparity, computing a disparity-dependent 2-D mapping from the original stereo pair to the synthesized views, and finally composing the synthesized views. In this paper, we compare different disparity-dependent mappings in terms of perceived shape distortion and alteration of the images, and we propose a hybrid mapping which does not distort depth and minimizes modifications of the image content.

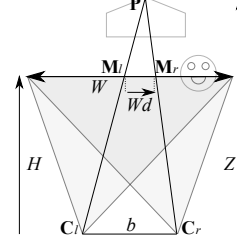
Index Terms— Image generation, Three-dimensional displays, Display human factors, Stereo vision

1. INTRODUCTION

As was shown in the early days of stereoscopic cinema, projecting a stereoscopic movie on different screen sizes and distances will produce different perceptions of depth [1, 2]. This implies that a stereoscopic film should be shot for a given display configuration, e.g. a movie theater room with a 10m wide screen placed at 15m, and displaying it with a different viewing geometry will distort depth, and may even cause eye divergence if the resulting on-screen disparities are bigger than the human interocular.

Let us study the distortions caused by given shooting and viewing geometries. The simple geometric parameters shown on Fig. 1 describe fully the stereoscopic setup, and their effects on shape perception are easier to understand than camera-based parameters used by previous approaches [3]. We assume that the stereoscopic movie is *rectified* and thus contains no vertical disparity, so that the convergence plane (where the disparity is zero) is vertical and parallel to the line joining the optical centers of the cameras.

The 3-D distortions in the perceived scene essentially come from different scene magnifications in the $X - Y$ directions, and in the Z direction. The ratio between depth magnification and width magnification is sometimes called *shape ratio* [2] or *depth reduction* [3], but we will use the term *roundness factor* in the remaining of our study. A low



Symbol	Camera	Display
C_l, C_r	camera optical center	eye optical center
P	physical scene point	perceived 3-D point
M_l, M_r	image points of P	screen points
b	camera interocular	eye interocular
H	convergence distance	screen distance
W	width of convergence plane	screen size
Z	real depth	perceived depth
d	left-to-right disparity (as a fraction of W)	

Fig. 1. Shooting and viewing geometries can be described using the same small set of parameters.

roundness factor will result in what is called the “cardboard effect”, and a rule of thumb used by stereographers is that it should never be below 0.2, or 20%.

Let b, W, H, Z be the stereoscopic camera parameters, and b', W', H', Z' be the viewing parameters, as described on Fig. 1. Triangles MPM' and CPC' are homothetic, consequently: $(Z - H)/Z = dW/b$.

It can easily be rewritten to get the image disparity d as a function of the real depth Z and vice-versa:

$$d = \frac{b}{W} \frac{Z - H}{Z}, \text{ or } Z = \frac{H}{1 - dW/b}. \quad (1)$$

For a fronto-parallel 3-D plane placed at distance Z , we can also compute the scale factor s from distances in the X and Y directions in that fronto-parallel plane to distances in the convergence plane: $s = H/Z$.

In the following, we will first consider how the 3-D shape is distorted by the viewing conditions. Then, we will investigate how new view synthesis can be used to avoid this effect, and we will consider three geometric transforms that can be used for that purpose: *baseline modification*, *viewpoint modification*, and *hybrid disparity remapping*. Baseline modification is the simplest method, but it cannot solve the problem of divergence at infinity while preserving the roundness factor. *Viewpoint modification* may cause major modifications of the original images due to changes in focal length. As a tradeoff between both methods, we propose *hybrid disparity remapping* which, while preserving depth perception and avoiding divergence, causes minimal deformations on the

original images.

2. VIEWING THE UNMODIFIED 3-D MOVIE

If the stereoscopic movie is viewed without modification, the horizontal disparity in the images, expressed as a fraction of image width, and the screen disparity, expressed as a fraction of screen width, are equal: $d' = d$. We can express the perceived depth Z' as a function of the disparity d :

$$Z' = \frac{H'}{1 - dW'/b'}. \quad (2)$$

Finally, eliminating the disparity d from eq. (1) and (2) gives the relation between real depth Z and perceived depth Z' :

$$Z' = \frac{H'}{1 - \frac{W'}{b'}(\frac{b}{W}\frac{Z-H}{Z})} \text{ or } Z = \frac{H}{1 - \frac{W}{b}(\frac{b'}{W'}\frac{Z'-H'}{Z'})} \quad (3)$$

Eye divergence happens when $Z' < 0$ or $d' > b'/W'$, and in general the objects that are at infinity in the real scene ($Z \rightarrow +\infty$) either cause divergence or are perceived at a finite depth.

We can also compute the *image scale ratio* σ' , which is how much an object placed at depth Z or at a disparity d seems to be enlarged ($\sigma' > 1$) or reduced ($\sigma' < 1$) in the X and Y directions with respect to objects that are in the convergence plane ($Z = H$):

$$\sigma' = \frac{s'}{s} = \frac{H'}{Z'} \frac{Z}{H} = \frac{1 - dW'/b'}{1 - dW/b}. \quad (4)$$

Of course, for on-screen objects ($d=0$), we have $\sigma' = 1$. Also note that the relation between Z and Z' is nonlinear, except if $W/b = W'/b'$, in which case $\sigma' = 1$ and the relation between Z and Z' simplifies to $Z' = ZH'/H$.

A small object of dimensions $\delta X \times \delta Z$ in the width and depth directions, placed at depth Z , is perceived as an object of dimensions $\delta X' \times \delta Z'$ at depth Z' , and the roundness factor ρ measures how much the object proportions are affected:

$$\rho = \frac{\partial Z'}{\partial Z} / \frac{\partial X'}{\partial X} = \frac{\partial Z'}{\partial Z} / \frac{W'/s'}{W/s} = \sigma' \frac{W}{W'} \frac{\partial Z'}{\partial Z} \quad (5)$$

In the screen plane ($Z = H$ and $Z' = H'$), the roundness factor simplifies to:

$$\rho_{\text{screen}} = \frac{W}{W'} \frac{\partial Z'}{\partial Z} \Big|_{(Z=H)} = \frac{b}{H} \frac{H'}{b'} \quad (6)$$

The roundness factor of an object in the screen plane is equal to 1 iff $b'/b = H'/H$, and adding the constraint that it must be equal to 1 everywhere (not only in the screen plane) leads to $b'/b = W'/W = H'/H$, which means that *the only shooting geometries that preserve the roundness factor everywhere are scaled versions of the viewing geometry*. Even if the viewing geometry is known, this imposes very hard constraints on the way the film must be shot, which may be impossible to follow in many situations (e.g. when filming wildlife or sports events). Besides, restricting the viewing geometry means that a film can only be projected on a given screen size W' placed at a given distance H' , since the human interocular b' is fixed.

3. NEW VIEW SYNTHESIS

When the shooting geometry is constrained, or when the same stereoscopic movie must be displayed with different viewing geometries, the shape distortions may be reduced by new view synthesis techniques [4, 5, 6], which usually involve three steps: computing the stereo disparity, computing a disparity-dependent 2-D mapping from the original stereo pair to the synthesized views, and finally composing the synthesized views from the original images, the disparity, and the disparity-dependent mapping.

For simplification purposes, we suppose that the position of the convergence plane within the scene and its width (i.e. the field size) are unchanged by this operation, but in practice the field size may be changed too. In the following, quantities referring to the synthesized geometry, i.e. the geometry of the synthesized image pair, are noted with double prime.

First, we may change the camera interocular b , in order to have $\rho = 1$ in the screen plane, i.e. $\frac{b'}{b} = \frac{H'}{H}$. To achieve this, we use *baseline modification*, a technique that generates a pair of new views *as if* they were taken by cameras placed at a specified position between the original camera positions. As shown in Fig. 2a, although the roundness factor of objects near the screen plane is well preserved, depth and size distortions are present and are what would be expected from the interpolated camera setup: far objects are heavily distorted both in size and depth, and divergence may happen at infinity.

If we also want to also change the distance to screen, we have to use *viewpoint modification*. It is a similar technique, where the synthesized viewpoint can be placed more freely. The problem is that we usually film with only two cameras, and large parts of the scene that would be visible in the synthesized viewpoint may not be visible in the original images, like the tree in Fig. 2b.

What we propose is a mixed technique between baseline modification and viewpoint modification, that preserves the global visibility of objects in the original viewpoints, but does not produce depth distortion or divergence: *hybrid disparity remapping*. In this method (Fig. 2c), we apply a nonlinear transfer function to the disparity function, so that perceived depth is proportional to real depth (the scale factor is computed at the convergence plane), and divergence may not happen, since objects that were at a given distance from the convergence plane on the original scene will be projected at the same distance, up to a fixed scale factor, and thus points at infinity are correctly displayed at infinity. However, since the apparent image size of the objects is not changed by hybrid disparity remapping, there will still be some kind of “puppet-theater” effect, and far-away objects may appear bigger in the image than they should be.

In order to compute the image transforms, let us examine the synthesized images of a constant-depth 3-D plane, and compare them with the original images of that same plane. With baseline modification and disparity remapping, we notice that there is only a horizontal shift (i.e. a disparity change) between the original and the synthesized images,

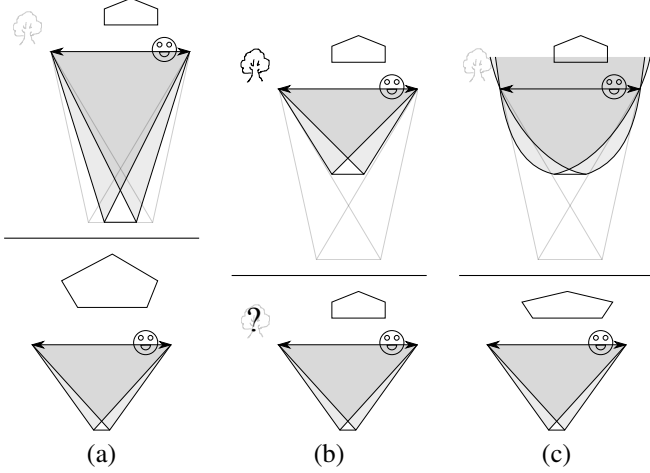


Fig. 2. Changing the shooting geometry in post-production (shooting geometry with synthesized geometry in dark gray, above viewing geometry with shape distortions): (a) baseline modification, (b) viewpoint modification, (c) hybrid disparity remapping. All methods can preserve the roundness factor of on-screen objects, at some cost for out-of-screen objects.

whereas with viewpoint modification a depth-dependent scale factor is also applied on the image of the constant-depth plane, because of the change in focal length. Thus, all three interpolations can be decomposed into a disparity-dependent scaling $\sigma''(d)$, and a disparity-dependent shift on the horizontal axis which depends on the synthesized disparity $d''(d)$.

Baseline modification only affects b'' , e.g. to get a roundness factor of $\rho = 1$ for on-screen objects: $b'' = b'H/H'$ (from eq. (6)). Since the disparity is proportional to the baseline b (eq. (1)), we obtain $d''(d) = db''/b$. Neither H , W or Z are changed, which implies that there is no disparity-dependent scaling ($\sigma''(d) = 1$). Points at infinity ($Z \rightarrow \infty$ or $d = b/W$) are mapped to $d''(b/W) = \frac{b'H}{WH'}$, and will cause divergence iff $\frac{b'H}{WH'} > \frac{b'}{W'}$.

Viewpoint modification does not change the size of the convergence plane ($W'' = W$), but the other parameters must be computed from the targetted viewing geometry: $b'' = b' \frac{W}{W'}$, $H'' = H' \frac{W}{W'}$. Since the scene objects must be at the same place with respect to the convergence plane in the shooting and the synthesized geometry, we have $Z'' - H'' = Z - H$, which can be rewritten using eq. (1) as:

$$d''(d) = \frac{Hb'd}{(HW' - H'W)d + H'b}. \quad (7)$$

The image scale ratio σ'' can be computed from eqs. (4), (1) and (7) as:

$$\sigma''(d) = \frac{H'b}{(HW' - H'W)d + H'b}. \quad (8)$$

There is no eye divergence with viewpoint modification, since points at infinity are mapped to $d''(b/W) = b'/W'$, i.e. $Z' \rightarrow \infty$. Note that if $H' = H$ and $W' = W$, we obtain the same formulas as for baseline modification.

Disparity remapping simply consists in taking the same syn-

thesized disparity as viewpoint modification (eq. (7)) - so that the depth is preserved and there is no eye divergence - and discarding the disparity-dependent image scaling: $\sigma''(d) = 1$.

From the synthesized disparity $d''(d)$ and image scale ratio $\sigma''(d)$ there are at least two options to synthesize an image pair. With *symmetric synthesis*, the original images have a symmetric role, and the synthesized views will have symmetric positions with respect to the mid-plane between the two optical centers C_l and C_r , as in Fig. 2. With *asymmetric synthesis*, one of the synthesized views is kept as close as possible to one of the original views, e.g. the left view. The reason for using asymmetric synthesis is that the perceived quality of a stereoscopic image pair is closer (and sometimes equal) to the quality of the best image in the pair [7]. Since new view synthesis is not perfect, synthesized images may contain artifacts and thus have a lower quality than the original images. We will see that with baseline modification or hybrid disparity remapping, one of the images in the pair can be left untouched, which would result in a perceived quality close to the original image quality.

In the following, (x_l, y) and (x_r, y) are respectively the left and right image coordinates, d_l is the left-to-right disparity, and d_r is the *opposite* of the right-to-left disparity.

Symmetric synthesis: The transfer function can be decomposed as:

1. map the original viewpoint to the cyclopean viewpoint (or the midpoint between the left and right camera positions);
2. compose with disparity-dependent scaling in the cyclopean view;
3. shift by the half synthesized disparity.

The resulting mappings are $(L, R, L'', R''$ denote respectively the left and right original images and the left and right synthesized images, w is the image width, and the mappings are from (x_l, y) in L and (x_r, y) in R :

$$L \rightarrow L'' : (x_c + (x_l + \frac{wd_l}{2} - x_c)\sigma''(d_l) - \frac{wd''(d_l)}{2}, y_c + (y - y_c)\sigma''(d_l))$$

$$L \rightarrow R'' : (x_c + (x_l + \frac{wd_l}{2} - x_c)\sigma''(d_l) + \frac{wd''(d_l)}{2}, y_c + (y - y_c)\sigma''(d_l))$$

$$R \rightarrow L'' : (x_c + (x_r - \frac{wd_r}{2} - x_c)\sigma''(d_r) - \frac{wd''(d_r)}{2}, y_c + (y - y_c)\sigma''(d_r))$$

$$R \rightarrow R'' : (x_c + (x_r - \frac{wd_r}{2} - x_c)\sigma''(d_r) + \frac{wd''(d_r)}{2}, y_c + (y - y_c)\sigma''(d_r))$$

The image weights, which measure how close each synthesized image is to each original image, and may be used as blending factors, can be computed as follow (we call α_A^B the weight of image $A \in \{L, R\}$ in synthesized image $B \in \{L'', R''\}$):

$$\alpha_{L''}^{L'}(d_l) = 1 - \frac{d_l - d''(d_l)}{2d_l} \quad \alpha_{L''}^{R'}(d_l) = \frac{d_l - d''(d_l)}{2d_l} \quad (9)$$

$$\alpha_{R''}^{L'}(d_r) = 1 - \frac{d_r - d''(d_r)}{2d_r} \quad \alpha_{R''}^{R'}(d_r) = \frac{d_r - d''(d_r)}{2d_r} \quad (10)$$

We notice that in the baseline modification case, those weights are constants, since $d''(d) = db''/b$.

Asymmetric Synthesis: With asymmetric synthesis, only the left image is used to compute the left synthesized image. The

mappings can be decomposed as:

1. apply disparity-dependent scaling in each view;
2. shift the right image by the difference between the synthesized disparity and the original disparity.

The resulting mappings are:

$$\begin{aligned} L &\rightarrow L'' : (x_c + (x_l - x_c)\sigma''(d_l), y_c + (y - y_c)\sigma''(d_l)) \\ L &\rightarrow R'' : (x_c + (x_l + wd_l - x_c)\sigma''(d_l) + w(d''(d_l) - d_l), y_c + (y - y_c)\sigma''(d_l)) \\ R &\rightarrow R'' : (x_c + (x_r - x_c)\sigma''(d_r) + w(d''(d_r) - d_r), y_c + (y - y_c)\sigma''(d_r)) \end{aligned}$$

and the image weights in the right image are:

$$\alpha_L^{R''}(d_l) = d''(d_l)/d_l \quad \alpha_R^{R''}(d_r) = 1 - d''(d_r)/d_r \quad (11)$$

With baseline modification and hybrid disparity remapping, $\sigma'' = 1$, so that the left image is not modified.

4. RESULTS

In order to show how new view synthesis is affected by the choice of the disparity-dependent mapping, we show partial results obtained from ground-truth disparity data, and we only show the left image mapped onto the left synthesized image, using symmetric synthesis. That way, we can visualize areas where no original image data is available (in red). Results were obtained with $W = 1m$, $H = 5m$, $b = 17.5cm$, $W' = 5m$, $H' = 15m$, $b' = 6.5cm$, and the convergence plane is at the depth of the statue's nose ($d_0 = 22px$ in the original data). The results show that with viewpoint modification, large areas contain no original information and would probably have to be inpainted. The results of hybrid disparity remapping and baseline modification look similar, but viewing the stereo pair obtained from baseline modification would cause eye divergence at infinity, whereas hybrid disparity remapping reproduces depth faithfully and does not cause divergence.

5. CONCLUSION

We presented three different disparity-dependent 2-D mappings for new view synthesis applied to stereoscopic cinema. All three mappings preserve the 3-D *roundness factor* of objects that are displayed near the screen depth, but they have different properties for objects that are displayed closer or farther than the screen: *baseline modification* does not solve the problem of eye divergence at infinity, *viewpoint modification* may generate very different viewpoints, where original image data is missing, and our method, *hybrid disparity remapping* is a good tradeoff, since it preserves depth, does not cause eye divergence, and generates images that are close to the original images, but still distorts the apparent image size of out-of-screen objects. Of course, other mappings may be used, in order to get specific artistic effects, but they should probably take *hybrid disparity remapping* as a starting point, since it seems to be the best compromise among roundness-preserving mappings.

Given a disparity mapping, we showed that new view synthesis can either be symmetric, i.e. both images have the same

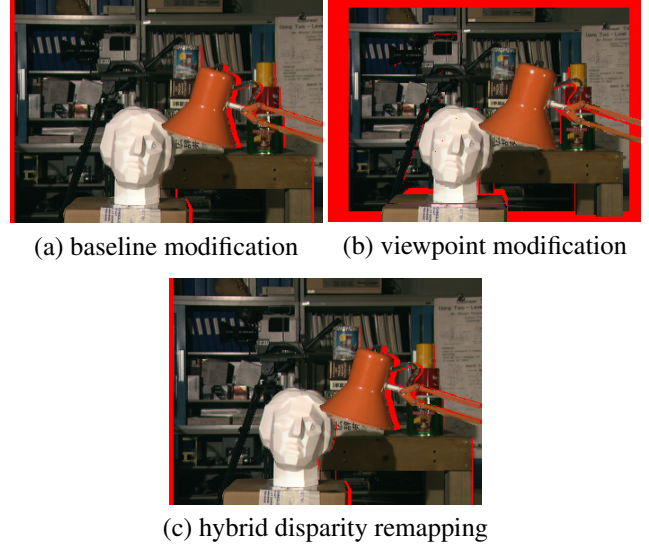


Fig. 3. The left image mapped onto the left synthesized image: in red are areas where there is no original image data.

role, or asymmetric. Asymmetric synthesis has the advantage that one of the images in the stereo pair can be left untouched, which should result in a better perceived quality of the stereo pair. Further experiments should be led on the perception of asymmetrically-synthesized stereoscopic sequences to get a better idea of the allowed amount of degradation in one of the images.

6. REFERENCES

- [1] Frédéric Devernay and Paul Beardsley, “Stereoscopic cinema,” in *Image and Geometry Processing for 3-D Cinematography*, Rémi Ronfard and Gabriel Taubin, Eds. Springer-Verlag, 2010.
- [2] Raymond Spottiswoode, N. L. Spottiswoode, and Charles Smith, “Basic principles of the three-dimensional film,” *SMPTE Journal*, vol. 59, pp. 249–286, Oct. 1952.
- [3] H. Yamanoue, M. Okui, and F. Okano, “Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, no. 6, pp. 744–752, June 2006.
- [4] C. Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski, “High-quality video view interpolation using a layered representation,” in *Proc. ACM SIGGRAPH*, New York, NY, USA, 2004, vol. 23, pp. 600–608, ACM.
- [5] A. Criminisi, A. Blake, C. Rother, J. Shotton, and P. H. Torr, “Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming,” *Int. J. Comput. Vision*, vol. 71, no. 1, pp. 89–110, 2007.
- [6] Sammy Rogmans, Jiangbo Lu, Philippe Bekaert, and Gauthier Lafruit, “Real-time stereo-based view synthesis algorithms: A unified framework and evaluation on commodity GPUs,” *Signal Processing: Image Communication*, vol. 24, no. 1-2, pp. 49–64, 2009, Special issue on advances in three-dimensional television and video.
- [7] Pieter Seuntjens, Lydia Meesters, and Wijnand Ijsselstein, “Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation,” *ACM Trans. Appl. Percept.*, vol. 3, no. 2, pp. 95–109, Apr. 2009.