Adapting stereoscopic movies to the viewing conditions using depth-preserving and artifact-free novel view synthesis

Frédéric Devernay, Sylvain Duchêne, Adrian Ramos-Peon

Adapting stereoscopic movies to the viewing conditions using depth-preserving and artifact-free novel view synthesis

Frédéric Devernay, Sylvain Duchêne and Adrian Ramos-Peon

INRIA Grenoble - Rhône-Alpes, France

ABSTRACT

The 3D 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 novel view synthesis techniques. They usually involve three steps: computing the stereo disparity, computing a disparity-dependent 2D mapping from the original stereo pair to the synthesized views, and finally composing the synthesized views. In this paper, we focus on the second and third step: we examine how to generate new views so that the perceived depth is similar to the original scene depth, and we propose a method to detect and reduce artifacts in the third and last step, these artifacts being created by errors contained in the disparity from the first step.

Keywords: Stereoscopic cinema, 3DTV, Stereoscopic display size, Novel view synthesis, View interpolation.

1. INTRODUCTION

The 3D 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 novel view synthesis techniques. They usually involve three steps: computing the stereo disparity, computing a disparity-dependent 2D mapping from the original stereo pair to the synthesized views, and finally composing the synthesized views.

Stereo disparity computation itself is a very active research topic in computer vision, and recent advances showed that in most cases very a accurate disparity map can be computed in a reasonable time (even sometimes at video-rate). However, difficult situations such as reduced depth-of-field, low-texture areas, depth discontinuities, repetitive patterns, transparencies or specular reflections are still very challenging and cause local errors in most disparity computation methods, which result in 2D or 3D artifacts in synthesized novel views.

In the first part[1] of this paper, we focus on the second step of novel view synthesis. First, we compute how the perceived 3D geometry is affected by the viewing conditions, and consider three disparity-dependent 2D mappings to adapt the stereoscopic movie to the viewing conditions, so that the perceived geometry is consistent with the original scene. The traditional baseline modification method consists in virtually changing the baseline between the two cameras, but whereas it may preserve the perceived 3D shape of objects that are close to the screen plane, severe depth distortions may appear on off-screen objects, and eye divergence may happen on distant objects. Viewpoint modification is another mapping which preserves the 3D shape of all objects in the scene, but may cause many large occluded areas (called disocclusions) to become visible in the novel views. In these areas, image content has to be recreated by complicated algorithms such as stereoscopic inpainting, and thus many artefacts may appear. We finally introduce hybrid disparity remapping, a new technique which preserves depth and causes no divergence, like viewpoint modification, but preserves image content and causes few disocclusions like baseline modification.

This work was done within the 3DLive project supported by the French Ministry of Industry `http://3dlive-project.com/`

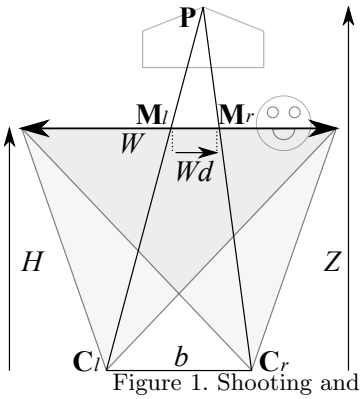
F. Devernay, S. Duchêne and A. Ramos-Peon, "Adapting stereoscopic movies to the viewing conditions using depth-preserving and artifact-free novel view synthesis", Stereoscopic Displays & Applications XXII, Woods, Holliman, Dogson (eds.), *Proc. SPIE* 7863, paper 7863-1, 2011. `http://dx.doi.org/10.1117/12.??????`

In the second part,[2] we discuss the last step of novel view synthesis: composing the synthesized views. Unfortunately, since no known disparity computation method gives perfect results in all situations, the results will most probably contain errors, which result in 2D or 3D artifacts in the synthesized stereoscopic movie. We show that in the case of baseline modification or hybrid disparity remapping, only one view needs to be synthesized. Previous work on asymmetric compression of stereoscopic movies showed that if one view is slightly blurred, the perceived quality of the stereo pair is close to this of the non-blurred view. We thus propose to add a post-processing phase where these artifacts are detected and blurred in the synthesized view, while keeping the perceived quality of the stereoscopic movie close to the original.

2. DEPTH-PRESERVING NOVEL VIEW SYNTHESIS

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.[3,4] 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.[5] 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.



Symbol	Camera	Display
$\mathbf{C}_l$, $\mathbf{C}_r$	camera optical center	eye optical center
$\mathbf{P}$	physical scene point	perceived 3-D point
$\mathbf{M}_l$, $\mathbf{M}_r$	image points of $\mathbf{P}$	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)	

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

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*[4] or *depth reduction*,[5] but we will use the term *roundness factor* in the remaining of our study. A low 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 $\mathbf{MPM'}$ and $\mathbf{CPC'}$ 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:

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

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

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 method, we propose *hybrid disparity remapping* which, while preserving depth perception and avoiding divergence, causes minimal deformations on the original images.

2.1 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:

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

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

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

Eye divergence happens when $Z' < 0$ or $d' > b'/W'$, and in general the objects that are at infinity in the real scene ($Z \to +\infty$) 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 ($Z = H$):

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

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 $Z'=ZH'/H$.

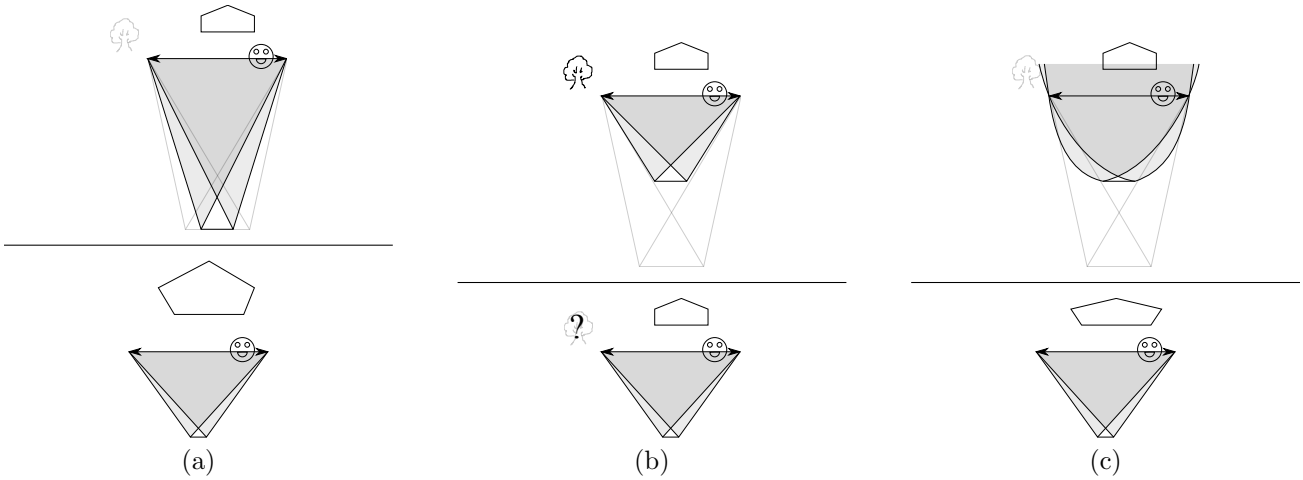
A small object of dimensions $\delta X \times \delta Z$ in the width and depth directions, placed at depth Z, is perceived as an object of dimensions $\delta X' \times \delta Z'$ at depth Z', and the roundness factor ρ measures how much the object proportions are affected:

$$\rho = \frac{\partial Z'}{\partial Z} \Big/ \frac{\partial X'}{\partial X} = \frac{\partial Z'}{\partial Z} \Big/ \frac{W'/s'}{W/s} = \sigma' \frac{W}{W'}\frac{\partial Z'}{\partial Z} \tag{5}$$

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

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

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.



(a) (b) (c)

Figure 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.

2.2 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,[6–12] 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 refering 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, 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 $d''(\frac{b}{W})=\frac{b'H}{WH'}$, 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:

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

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

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

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.

Hybrid disparity remapping simply consists in taking the same synthesized 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 $\mathbf{C}_l$ and $\mathbf{C}_r$, 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.[13] 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):

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

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:

$$L \to L'' : (x_c + (x_l - x_c)\sigma''(d_l), y_c + (y - y_c)\sigma''(d_l))$$
$$L \to R'' : (x_c + (x_l + wd_l - x_c)\sigma''(d_l) + w(d''(d_l) - d_l), y_c + (y - y_c)\sigma''(d_l))$$
$$R \to R'' : (x_c + (x_r - x_c)\sigma''(d_r) + w(d''(d_r) - d_r), y_c + (y - y_c)\sigma''(d_r))$$

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

2.3 Example

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 ($d_0 = 22$px 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.

(a) baseline modification (b) viewpoint modification (c) hybrid disparity remapping

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

2.4 Avoiding the vergence-accomodation conflict

When viewing a stereoscopic movie, the distance of accomodation differs from the distance of convergence, which is the distance to the perceived object. For example, for a 3DTV screen placed at 3m, the depth-of-field goes from 1.9m to 7.5m, and objects that are displayed outside of this range will cause visual fatigue, caused by the vergence-accomodation conflict. This means that the in-focus displayed objects should have disparities between $6.5 \times (1.9-3)/1.9 = -3.8$cm and $6.5 \times 3/7.5 = 2.6$cm. If the depth-of-field can be reduced to match these values, the vergence-accomodation conflicts can be attenuated,[14] but this would also destroy image content (a far-away scenery would be completely blurred that way).

Depth-preserving novel-view synthesis can also be tweaked to remain within these limits: we can keep the $\rho_{\text{screen}} = 1$ constraint for the objects in the convergence plane, but the screen size W' will be computed so that the farthest in-focus objects have a disparity of 2.6cm. Since the on-screen roundness factor does not depend on the screen size, we keep a screen distance of 3m, and the resulting synthesized stereoscopic movie will be a good compromise between depth preservation of on-screen objects and respect of the vergence-accomodation constraints. In some cases, especially when the farthest in-focus objects are at infinity, it may also be a good idea to reduce the on-screen roundness factor by using a larger screen distance for novel-view synthesis ($H' = 6$m will result in $\rho_{\text{screen}} = 0.5$ for a real viewing distance of 3m), in ordre to reduce nonlinear depth distortions in the Z direction.

3. ARTIFACTS DETECTION AND REMOVAL

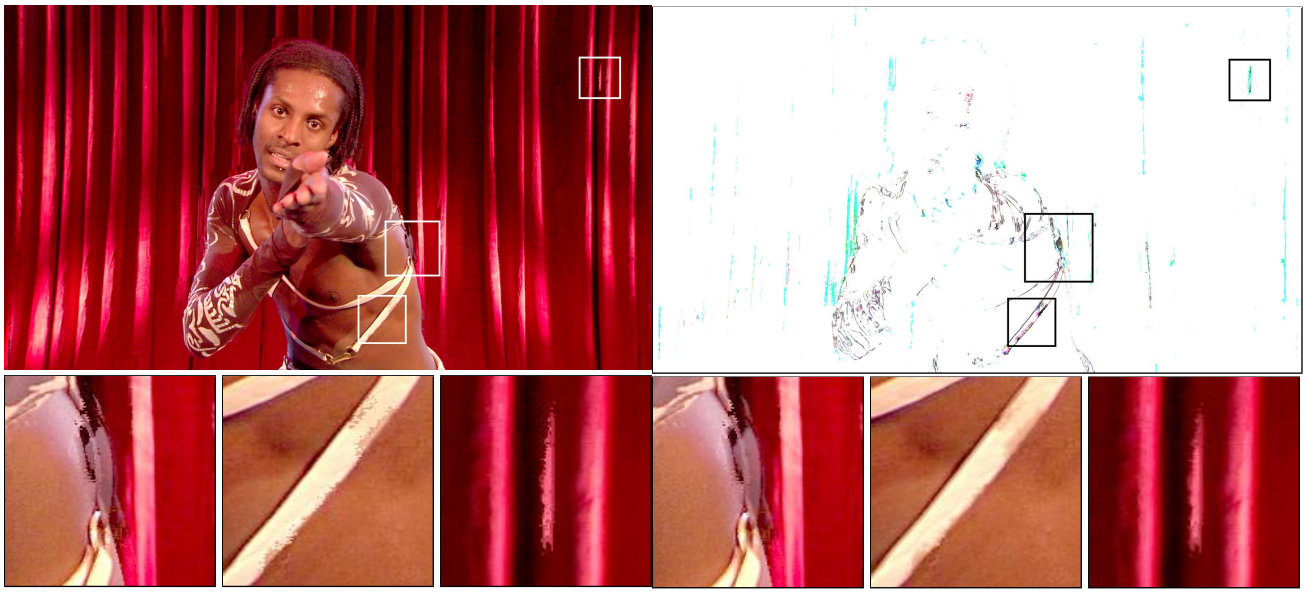


Figure 4. Left: synthesized novel view with zoom on three artifacts. Right: confidence map used to detect artifacts and results of artifact removal.

Novel view synthesis from stereoscopic movies wer reviewed by Rogmans *et al.*,[8] who noticed that they essentially consist of two steps: first, a stereo correspondence module computes the stereoscopic disparity between the two views, and second, a view synthesis module generates the new views, given the results of the first module and the parameters of the synthesized cameras. The main consequence is that any error in the first module will generate artifacts in the generated views. These can either be 2D artifacts, which appear only on one view and may disrupt the perceived scene quality and understanding, or even worse: 3D artifacts, that may appear as floating bits in 3D and look very unnatural.

We thus propose to add a third module that will detect artifacts, and remove them by smoothing them out. The key idea is that stereoscopic novel view synthesis can be done in an asymmetric way. As noted by Seuntiens

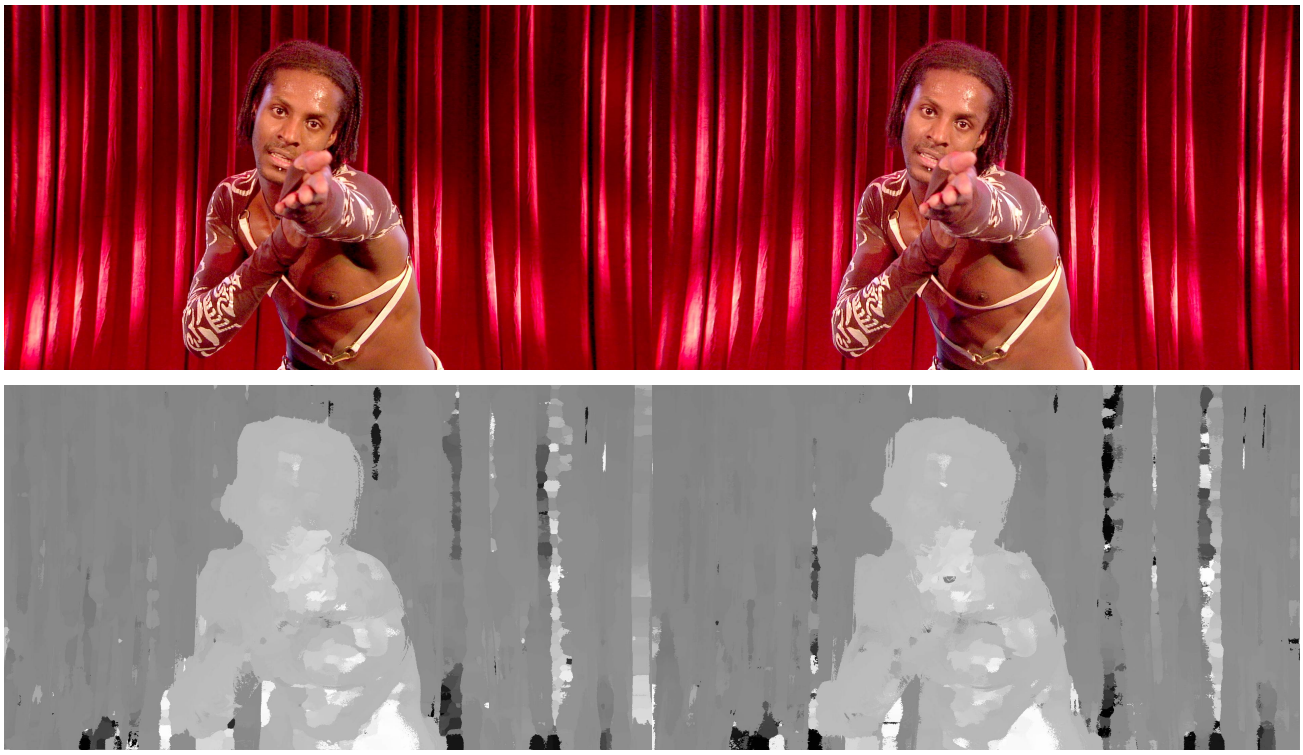


Figure 5. Top row: the left and right rectified images from the original stereo pair (pictures courtesy of Binocle). Bottom row: the left and right disparity maps produced from these images by a basic method, obviously containing many matching errors.

et al.,[13] if one of the views is close to or equal to the original image, the other view can be slightly degraded without any negative impact on the perceived quality of the stereoscopic movie, and eye dominance has no effect on the quality. We thus propose to use *asymmetric novel view synthesis*, where the left view is the original image, and only the right view is synthesized. Consequently, artifacts are only present in the right view, and we propose to detect and remove them by smoothing.[2] In these very small modified areas of the stereoscopic image pair, the visual system will use the left view combined with 3D cues other than stereopsis to reconstruct the proper 3D geometry of the scene.

We can assume that the original images were rectified, so that there is no vertical disparity between the two images: epipolar lines are horizontal, and a point (x_l, y) in the left image corresponds to a point at the same y coordinate (x_r, y) in the right image. The 3D information about the part of the scene that is visible in the stereo pair is fully described by the camera parameters, and the disparity maps that describe the mapping between points in the two images.

Let $I_l(x_l, y)$, $I_r(x_r, y)$ be a pair of rectified images, and $d_l(x_l, y)$, $d_r(x_r, y)$ be respectively the left-to-right and right-to-left disparity maps: d_l maps a point (x_l, y) in the left image to the point $(x_l - d_l(x_l, y), y)$ in the right image, and d_r maps a point (x_r, y) in the right image to the point $(x_r + d_r(x_r, y), y)$ in the left image (signs are set so that the bigger the disparity, the closer the point). These two disparity maps may be produced by any method, and the semi-occluded areas, which have no correspondent in the other image, are supposed to be filled using some assumption on the 3D scene geometry. A stereoscopic pair of images and their corresponding disparity maps used in our examples are shown in Fig. 5.

In the synthesized view, each pixel may have a visible matching point in the left image and/or a visible correspondent in the right image. If the point is not visible in one of the original images, the mapping is undefined at that point. We call these mappings from the synthesized view to the original images *backward mappings*. We focus on asymmetric synthesis methods, where the left image in the output stereoscopic pair is

the original left image, and only the right image in the output stereoscopic pair in synthesized, so the viewpoint modification method cannot be used, and the backward mappings only have a horizontal component and can be represented by backward disparity maps.

Let d_i^l and d_i^r be the backward disparity maps from the interpolated viewpoint, respectively to the left and to the right original images (the subscript is the reference viewpoint, and the superscript is the destination image). d_i^l and d_i^r map each integer-coordinates point in the interpolated image I_i to a real-coordinates point in I_l and I_r, or to an undefined value if the point is not visible in the corresponding original image. From these backward disparity maps, we compute the interpolated image, usually by computing a weighted average of the colors taken from the original images. The weights can be computed from the absolute values of the backward disparities, e.g. $\alpha_i^l = |d_i^r|/(|d_i^l| + |d_i^r|)$ and $\alpha_i^r = 1 - \alpha_i^l$.

3.1 Artifacts detection

With the help of the backward disparity maps, we can get any kind of value for (almost) all pixels in the synthesized viewpoint, be it intensity, Laplacian or gradient, as long as it can be computed in the left and right images. To get $I_i(x_i, y)$, the pixel intensity at (x_i, y), we will begin by finding $I_i^l(x_i, y)$ and $I_i^r(x_i, y)$, that is, the intensities in the left and right images corresponding to each point (x_i, y) in the novel view. These are computed by linear interpolation of the intensities at position $d_i^l(x_i, y)$ of the values in I_l, and at $d_i^r(x_i, y)$ in I_r. The values of $d_i^l(x_i, y)$ and $d_i^r(x_i, y)$ might be invalid due to disocclusion, in which case the pixel value is either marked as invalid, or some hole-filling method is applied.[8]

Artifact detection works by building a confidence map over the whole interpolated image, where most pixels are marked with high confidence, and artifacts are marked with low confidence. Once an interpolated image has been generated, we want to create a confidence map, attributing a weight to each pixel to specify how certain we are about its correctness.

Having the original left and right viewpoints of a scene, and the interpolated viewpoint of the same scene, building this confidence map is based on the fact that we expect to find similar pixel intensities, gradients and Laplacians in all images (excluding occlusions), but at different locations due to the geometric mappings between these views. Using this observation, we are able to outline areas and edges which should not appear in the interpolated view. For instance, a gradient appearing in the synthesized view that does not exist in either of the two original views should suggest the presence of an artifact, and will be marked as a low confidence zone in our confidence map. We thus use the backward mappings d_α^l and d_α^r to compare the not only the intensities, bu also the gradients and Laplacians of the interpolated view with the left and right views*.

The artifacts that appear in the synthesized view are mainly composed of high frequency components of the image, thus the Laplacian differences should give a good hint on where the artifacts are located, but only detects the contour of the artifacts, not their inner region. Intensity or gradient differences, on the other hand, may appear at many different places which are not actual artifacts, such as large specular reflections, or intensity differences due to the difference in illumination, but they also cover the inner regions of actual artifacts. We thus want to detect artifacts as areas which are surrounded by high Laplacian differences and inside which the intensity or gradient difference with the original images is high. We start by dilating the Laplacian using a small structuring element (typically a 3×3 square), so that not only the borders are detected, but also more of the inner (and outer) region of the artifact. Then, to remove the regions outside the actual artifacts (introduced by the dilation), we multiply this dilated map by the intensity difference map. This multiplication partly alleviates the specularity problem, since the regions detected as "uncertain" by the intensity map alone are now compared using the Laplacian, and if there are no discrepancies in the Laplacian differences, this area will be marked as being correct in the resulting confidence map. To further avoid incorrect detections introduced by the intensity differences, we decide to discard the weakest values. To do so, we set a threshold so that at most 5% of the image will have a non-zero value in the confidence map. This prevents overall image blurring in subsequent treatment of the interpolated image. The confidence map obtained in this way on the sample stereo pair and synthesized

*Theoretically, warped derivatives (gradients and Laplacian) should be composed with the derivatives of the mappings, but we make the assumption that the scene surface is locally fronto-parallel and ignore these, because the mappings derivatives contain too much noise

view is shown in Figure 4 (middle row) and details are shown in Figure 6. As can be seen, larger artifacts are indeed well detected.

Figure 6. Zoom of the color inverted confidence map on the artifacts from Fig. 4 (white corresponds to zero in the confidence map, representing correct values).

3.1.1 Artifact removal by anisotropic blurring

To try to remove as much as possible of the detected artifacts, we use the anisotropic diffusion equation, as used by Perona-Malik:[15]

$$\frac{\partial I}{\partial t} = \nabla \cdot (c(x,y,t)\nabla I) = c(x,y,t)\Delta I + \nabla c \cdot \nabla I \tag{9}$$

where $c(x,y,t)$ are the conduction coefficients, which guide the smoothing of the image. While Perona and Malik were interested in finding the coefficients to smooth the image only within regions of slowly-varying color, and not across boundaries, we already know where we want the diffusion to occur. The confidence map provides these space variant coefficients, that cause the detected artifacts to be smoothed out, while the rest of the image is left untouched. The values from the confidence map are normalized so that the resulting coefficients are between zero and one. In this case, the Perona-Malik equation is proved to converge, and the numerical scheme used to implement the proposed smoothing is very simple.[15] In our example, the anisotropic blurring is performed with a time step of $\Delta t = 0.25$, and 20 iterations are computed.

The right synthesized view is shown on Fig. 7, and zoom on details is available on Fig. 4. Small and medium artifacts were detected and removed by our algorithm, but some of the bigger artifacts are still present. We notice for example on Fig. 4 that the "curtain" artifact was not completely removed because there is a very large matching error in the original disparity maps, due to a repetitive pattern with slight occlusions (the curtain folds), and part of the resulting artifact is consistent with the original images and disparity maps, as can be seen in the confidence map (Fig. 6). It proves that the disparity maps still have to be of an acceptable quality in order to remove properly all the artifacts, and the final quality of the stereo pair still depends on the quality of the stereo correspondence module, although in lesser proportions than if this artifact removal module is not present: a state-of-the art stereo correspondence methods will produce less and smaller artifacts which will be easily removed by the proposed method (but the artifacts would be almost unnoticeable on a monoscopic image, although they still appear when viewed in 3D).

Some of the natural artifacts that should be present in the synthesized image, such as specular reflections in the eyes, were also smoothed a little, but the impact on the resulting perceived quality of the stereoscopic pair is not important, since the left image still has these natural artifacts (specular reflections do not follow the epipolar constraint, and are thus rarely matched between the two views, even in the human visual system, although they still bring a curvature cue on the local surface geometry).

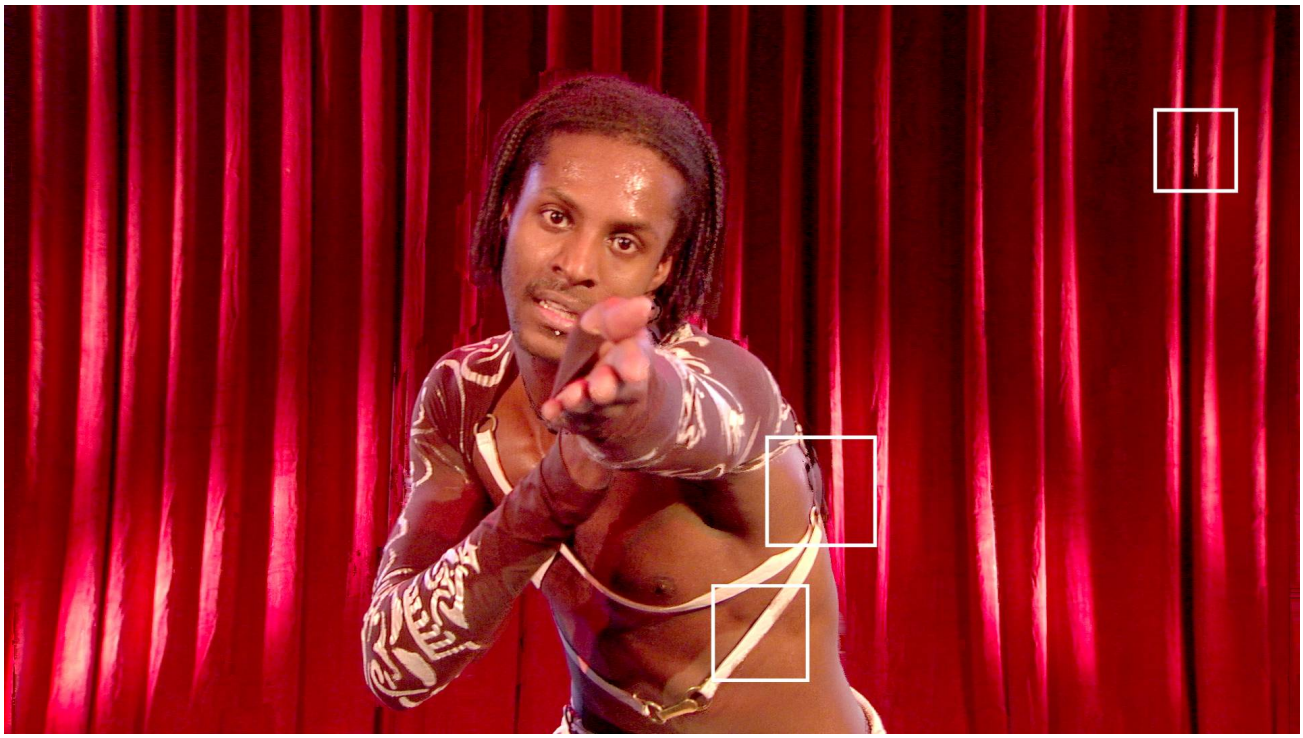


Figure 7. The synthesized right image, after artifact removal (compare with top row of Fig. 4).

4. CONCLUSION AND FUTURE WORK

We presented a complete method for artifact-free depth-preserving novel view synthesis for stereoscopic movies. We first showed that, given the fact that shooting and viewing geometries are usually different, a novel view synthesis has to be applied in order to preserve the depth proportions in the scene and the roundness factor. We proposed a novel interpolation function, *hybrid disparity remapping*, which 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. Furthermore, it can be adapted to deal with the vergence-accomodation conflict.

Novel view synthesis methods for stereoscopic video usually rely on two algorithmic modules which are applied in sequence to each stereoscopic pair in the movie:[8] a stereo correspondence module and a view synthesis module. Unfortunately, in difficult situations such as occlusions, repetitive patterns, specular reflections, low texture, optical blur, or motion blur, the stereoscopic correspondence module produces errors which appear as artifacts in the final synthesized stereo pair. We showed that in the case of hybrid disparity remapping, only one of the views had to be synthesized, so that artifacts are only present in one view. We detects artifacts in the synthesized view by producing a confidence map, and then smooth out these artifacts by anisotropic diffusion based on the Perona-Malik equation.[15]

The results show that this method removes small artifacts from the synthesized view. However, large artifact that are consistent both with the original images and the disparity maps may remain after this process, so the quality of the stereo correspondence module is still crucial for artifact-free novel view synthesis. Since these preliminary results are promising, we intend to work on the validation of this method by a psycho-visual study involving several viewers, in order to evaluate quantitatively the quality improvements brought by the application of this artifact removal module. We also plan to work on integrating temporal consistency by the combined use of disparity maps and optical flow maps, in order to reduce the appearance of flickering artifacts in stereoscopic movies, which are probably the most disturbing spatial artifacts for the movie viewer.

REFERENCES

[1] Devernay, F. and Duchêne, S., "New view synthesis for stereo cinema by hybrid disparity remapping," in [*International Conference on Image Processing (ICIP)*], 5 –8 (Sept. 2010).

[2] Devernay, F. and Peon, A. R., "Novel view synthesis for stereoscopic cinema: detecting and removing artifacts," in [*Proceedings of the 1st international workshop on 3D video processing*], *3DVP '10*, 25–30, ACM, New York, NY, USA (2010).

[3] Devernay, F. and Beardsley, P., "Stereoscopic cinema," in [*Image and Geometry Processing for 3-D Cinematography*], Ronfard, R. and Taubin, G., eds., Springer-Verlag (2010).

[4] Spottiswoode, R., Spottiswoode, N. L., and Smith, C., "Basic principles of the three-dimensional film," *SMPTE Journal* **59**, 249–286 (Oct. 1952).

[5] Yamanoue, H., Okui, M., and Okano, F., "Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images," *IEEE Trans. on Circuits and Systems for Video Technology* **16**, 744–752 (June 2006).

[6] Zitnick, C. L., Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R., "High-quality video view interpolation using a layered representation," in [*Proc. ACM SIGGRAPH*], *ACM Trans. Graph.* **23**(3), 600–608, ACM, New York, NY, USA (2004).

[7] Criminisi, A., Blake, A., Rother, C., Shotton, J., and Torr, P. H., "Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming," *Int. J. Comput. Vision* **71**(1), 89–110 (2007).

[8] Rogmans, S., Lu, J., Bekaert, P., and Lafruit, G., "Real-time stereo-based view synthesis algorithms: A unified framework and evaluation on commodity GPUs," *Signal Processing: Image Communication* **24**(1-2), 49–64 (2009). Special issue on advances in three-dimensional television and video.

[9] Koppal, S. J., Zitnick, C. L., Cohen, M., Kang, S. B., Ressler, B., and Colburn, A., "A viewer-centric editor for stereoscopic cinema," *IEEE Computer Graphics and Applications* **PP**(99), 1 (2010).

[10] Farin, D., Morvan, Y., and de With, P. H. N., "View interpolation along a chain of weakly calibrated cameras," in [*IEEE Workshop on Content Generation and Coding for 3D-Television*], (2006).

[11] Kilner, J., Starck, J., and Hilton, A., "A comparative study of free-viewpoint video techniques for sports events," in [*Proc. 3rd European Conference on Visual Media Production*], 87–96 (2006).

[12] Woodford, O., Reid, I. D., Torr, P. H. S., and Fitzgibbon, A. W., "On new view synthesis using multiview stereo," in [*Proceedings of the 18th British Machine Vision Conference*], **2**, 1120–1129 (2007).

[13] Seuntiens, P., Meesters, L., and Ijsselsteijn, W., "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Trans. Appl. Percept.* **3**, 95–109 (Apr. 2009).

[14] Ukai, K. and Howarth, P. A., "Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations," *Displays* **29**, 106–116 (Mar. 2007).

[15] Perona, P. and Malik, J., "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**, 629–639 (1990).