

Stereoscopic Cinema

Frédéric Devernay, Paul Beardsley

▶ **To cite this version:**

Frédéric Devernay, Paul Beardsley. Stereoscopic Cinema. Ronfard, Rémi and Taubin, Gabriel. Image and Geometry Processing for 3-D Cinematography, 5, Springer Berlin Heidelberg, pp.11-51, 2010, Geometry and Computing, 978-3-642-12392-4. 10.1007/978-3-642-12392-4 . inria-00509217v2

HAL Id: inria-00509217

https://inria.hal.science/inria-00509217v2

Submitted on 4 Jun 2015

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.



Distributed under a Creative Commons Attribution - ShareAlike| 4.0 International License

Stereoscopic Cinema*

Frédéric Devernay and Paul Beardsley

Abstract Stereoscopic cinema has seen a surge of activity in recent years, and for the first time all of the major Hollywood studios released 3-D movies in 2009. This is happening alongside the adoption of 3-D technology for sports broadcasting, and the arrival of 3-D TVs for the home. Two previous attempts to introduce 3-D cinema in the 1950s and the 1980s failed because the contemporary technology was immature and resulted in viewer discomfort. But current technologies – such as accurately-adjustable 3-D camera rigs with onboard computers to automatically inform a camera operator of inappropriate stereoscopic shots, digital processing for post-shooting rectification of the 3-D imagery, digital projectors for accurate positioning of the two stereo projections on the cinema screen, and polarized silver screens to reduce cross-talk between the viewers left- and right-eyes – mean that the viewer experience is at a much higher level of quality than in the past. Even so, creation of stereoscopic cinema is an open, active research area, and there are many challenges from acquisition to post-production to automatic adaptation for different-sized display. This chapter describes the current state-of-the-art in stereoscopic cinema, and directions of future work.

Frédéric Devernay
INRIA Grenoble - Rhône-Alpes, France e-mail: frederic.devernay@inria.fr

Paul Beardsley
Disney Research, Zürich, e-mail: pab@disneyresearch.com

* This was published as chapter 2 in Rémi Ronfard and Gabriel Taubin, editors, Image and Geometry Processing for 3-D Cinematography, volume 5 of Geometry and Computing, pages 1151. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-12392-4. doi: 10.1007/978-3-642-12392-4_2. The original publication is available at www.springerlink.com.

1 Introduction

1.1 Stereoscopic Cinema, 3-D Cinema, and Others

Stereoscopic cinema is the art of making stereoscopic films or motion pictures. In stereoscopic films, depth perception is enhanced by having different images for the left and right human eye, so that objects present in the film are perceived by the spectator at different depths. The visual perception process that reconstructs the 3-D depth and shape of objects from the left and right images is called stereopsis.

Stereoscopic cinema is also frequently called 3-D cinema[2], but in this work we prefer using the term stereoscopic, as 3-D cinema may also refer to other novel methods for producing moving pictures:

- free-viewpoint video, where videos are captured from a large number of cameras and combined into 4-D data that can be used to render a new video as seen from an arbitrary viewpoint in space;
- 3-D geometry reconstructed from multiple viewpoints and rendered from an arbitrary viewpoint.

Smolic et al [55] did a large review on the subjects of 3-D and free-viewpoint video, and other chapters in the present book also deal with building and using these moving pictures representations.

In this chapter, we will limit ourselves to movies that are made using exactly two cameras placed in a stereoscopic configuration. These movies are usually seen through a 3-D display device which can present two different images to the two human eyes (see Sexton and Surman [52] for a review on 3-D displays). These display devices sometimes take as input more than two images, especially glasses-free 3-D displays where a number of viewpoints are displayed simultaneously but only two are seen by a human spectator. These viewpoints may be generated from stereoscopic film, using techniques that will be described later in this chapter, or from other 3-D representations of a scene.

Many scientific disciplines are related to stereoscopic cinema, and they will sometimes be referred to during this chapter. These include computer science (especially computer vision and computer graphics), signal processing, optics, optometry, ophthalmology, psychophysics, neurosciences, human factors, and display technologies.

1.2 A Brief History of Stereoscopic Cinema

The interest in stereoscopic cinema appeared at the same time as cinema itself, and the "prehistory" (from 1838 to 1922 when the first public projection was made [22, 84]) as well as "history" (from 1922 to 1952, when the first feature-length

[2] Stereoscopic cinema is even sometimes redundantly called Stereoscopic 3-D (or S3D) cinema.

commercial stereoscopic movie, *Bwana Devil*, was released to the public [84]) follow closely the development of cinematography. The history of the period from 1952 to 2004 can be retraced from the thoughtful conversations found in Zone [83].

In the early 50's, the television was causing a large decrease in the movie theater attendance, and stereoscopic cinema was seen as a method to get back this audience. This explains the first wave of commercial stereoscopic movies in the 50's, when Cinemascope also appeared as another television-killer. Unfortunately the quality of stereoscopic movies, both in terms of visual quality and in terms of cinematographic content, was not on par with 2-D movies in general, especially Cinemascope. Viewing a stereoscopic film at that time meant the promise of a headache, because stereoscopic filming techniques were not fully mastered, and the spectacular effects were overly used in these movies, reducing the importance of screenplay to close to nothing. One notable counter-example to the typical weak screenplay of stereoscopic movies was *Dial M for Murder* by Alfred Hitchcock, which was shot in 3-D, but incidentally had a much bigger success as a 2-D movie. Finally, Cinemascope won the war against TV, and the production of stereoscopic cinema declined.

The second stereoscopic cinema "wave", in the 80's, was formed both by large format (IMAX 3-D) stereoscopic movies, and by standard stereoscopic movies which tried again the recipe from the 50's (spectacular content but weak screenplay), with no more success than the first wave. There were a few high-quality movies, such as *Wings of Courage* by Jean-Jacques Annaud in 1995, which was the first IMAX 3-D fiction movie, made with the highest standards, but at a considerable cost. The IMAX-3D film camera rig is extremely heavy and difficult to operate [83]. Stereoscopic cinema wouldn't take off until digital processing finally brought the tools that were necessary to make stereoscopic cinema both easier to shoot and to watch.

As a matter of fact, the rebirth of stereoscopic cinema came from animation, which produced movies like *Chicken Little*, *Open Season* and *Meet the Robinsons*, which were digitally produced, and thus could be finely tuned to lower the strain on the spectator's visual system, and experiment new rules for making stereoscopic movies. Live-action movies produced using digital stereoscopic camera systems came afterwards, with movies such as *U2 3D*, and of course *Avatar*, which held promises of a stereoscopic cinema revival.

Stereoscopic cinema brought many new problems that were not addressed by traditional movie making. Many of these problems deal with geometric considerations: how to place the two cameras with respect to each other, where to place the actors, what camera parameters (focal length, depth of field, ...) should be used... As a matter of fact, many of these problems were somehow solved by experience, but opinions often diverged on the right solution to film in stereoscopy. The first theoretical essay on stereoscopic cinema was written by Spottiswoode et al [57]. The Spottiswoodes made a great effort to formalize the influence of the camera geometry on the 3-D perception by the audience, and also proposed a solution on the difficult problem of transitions between shots. Their scientific approach of stereoscopic cinema imposed very strict rules on moviemaking, and most of the stereoscopic moviemakers didn't think it was the right way to go [83].

Many years later, Lenny Lipton, who founded StereoGraphics and invented the CrystalEyes electronic shutter glasses, tried again to describe the scientific foundations of stereoscopic cinema [37]. His approach was more viewer-centric, and he focused more on how the human visual system perceives 3-D, and how it reacts to stereoscopic films projected on a movie screen. Although the resulting book contains many mathematical errors, and even forgets most of the previous finding by the Spottiswoodes [54], it remains one of the very few efforts in this domain.

The last notable effort at trying to formalize stereoscopic movie making was this of Mendiburu [42], who omitted maths, but explained with clear and simple drawings the complicated geometric effects that are involved in stereoscopic cinema. This book was instantaneously adopted as a reference by the moviemaking community, and is probably the best technical introduction to the domain.

1.3 Computer Vision, Computer Graphics, and Stereoscopic Cinema

The discussions in this chapter straddle Computer Vision, Computer Graphics, and Stereoscopic Cinema: Computer Vision techniques will be used to compute and locate the defects in the images taken by the stereoscopic camera, and Computer Graphics techniques will be used to correct these defects.

1.3.1 A Few Definitions

Each discipline has its own language and words. Before proceeding, let us define a few geometric elements that are useful in stereoscopic cinema [29]:

Interocular (also called Interaxial): (the term baseline is also widely used in computer vision, but not in stereoscopic cinema) The distance between the two eyes/cameras, or rather their optical centers. It is also used sometimes used to designate the segment joining the two optical centers. The average human interocular is 65mm, with large variations around this value.

Hyperstereo (or miniaturization): The process of filming with an interocular larger than 65mm (it can be up to a few dozen meters), with the consequence that the scene appears smaller when the stereoscopic movie is viewed by a human subject.

Hypostereo (or gigantism): The process of filming with an interocular smaller than 65mm, resulting in a "bigger than life" appearance when viewing the stereoscopic movie. It can be as small as 0mm.

Roundness factor: Suppose a sphere is filmed by a stereoscopic camera. When displaying it, the roundness factor is the ratio between its apparent depth and its apparent width. If it is lower than 1, the sphere appears as a flattened disc parallel to the image plane. If it is bigger than 1, the sphere appears as a spheroid elongated in the viewer's direction. We will see that the roundness factor depends on the object's position in space.

Disparity: The difference in position between the projections of a 3-D point in the left and right images, or the left and right retinas. In a standard stereo setup, the disparity is mostly horizontal (the corresponding points are aligned vertically), but vertical disparity may happen (and has to be corrected).

Screen plane: The position in space where the display projection surface is located (supposing the projection surface is planar).

Vergence, convergence, divergence : the angle formed by the the optical axis of the two eyes in binocular vision. The optical axis is the half 3-D line corresponding to the line-of-sight of the center of the fovea. It can be positive (convergence) or negative (divergence).

Plane of convergence: The vertical plane parallel to the screen plane containing the point that the two eyes are looking at. If it is in front of the screen plane, then the object being looked at appears in front of the screen. When using cameras, the plane of convergence is the zero-disparity plane (it is really a plane if images are rectified, as will be seen later).

Proscenium arch (also called stereoscopic window and floating window): (Fig. 5) The perceived depth of the screen borders. If the left and right borders of the left and right images do not coincide on the screen, the proscenium arch is not in the screen plane. It is also called the stereoscopic window, since the 3D-scene looks as if it were seen through that 3-D window. As will be explained later, objects closer than the proscenium arch should not touch the left or right side of the arch.

1.3.2 Stereo-Specific Processes

The stereoscopic movie production pipeline shares a lot with standard 2-D movie production [42]. Computer vision and computer graphics tools used in the process can sometimes be used with no or little modification. For example, matchmoving (which is more often called *Structure from Motion* or *SfM* in Computer Vision) can take into account the fact that two cameras are taking the same scene, and may use the additional constraint that the two camera positions are fixed with respect to each other.

Many processes, though, are specific to stereoscopic cinema production and post-production, and cannot be found in 2-D movie production:

- *Correcting geometric causes of visual fatigue* such as images misalignments, will be covered by the next two sections (2.5 and 5.1).
- *Color-balancing left and right images* is especially necessary when a half-silvered mirror is used to separate images for the left and right cameras and wide angle lenses are used: the transmission and reflection spectrum response of these mirrors depend on the incidence angle and have to be calibrated. This can be done using color calibration devices and will not be covered in this chapter.
- *Adapting the movie to the screen size (and distance)* is not as simple as scaling the left and right views: the stereoscopic display is not easily scalable like a 2-D display, because the human interocular is fixed and therefore does not “scale”

with the screen size or distance. A consequence is that the same stereoscopic movie displayed on screens of different sizes will probably give quite different 3-D effects (at least quantitatively). The adaptation can be done at the shooting stage (Sec. 3) or in post-production (Sec. 5). These processes can either be used to give the stereoscopic scene the most natural look possible (which usually means a roundness factor close to 1), or to "play" with 3-D effects, for example by changing the interocular or the position where the infinity plane appears.

- *Local 3-D changes (or 3-D touchup)* consist in editing the 3-D content of the stereoscopic scene. This usually means providing new interactive editing tools that work both on the images and on the disparity map. These tools usually share a lot with colorizing tools, since they involve cutting-up objects, tracking them, and changing their depth (instead of color). This is beyond the scope of this chapter.
- *Playing with the depth of field* is sometimes necessary, especially when adapting the 3-D scene to a given screen distance: the depth of field should be consistent with the distance to the screen, in order to minimize vergence-accomodation conflicts (Sec. 5.4)
- *Changing the proscenium arch* is sometimes necessary, because objects in front of the screen may cross the screen borders and become inconsistent with the stereoscopic window (Sec 5.5).
- *3-D compositing (with real or CG scenes)* should be easier in stereoscopic cinema, since it has 3-D content already. However, there are some additional difficulties: 2-D movies mainly have to deal with positioning the composited objects within the scene and dealing with occlusion masks. In 3-D, the composited scene must also be consistent between the two eyes, and its 3-D shape must be consistent with the original stereoscopic scene (Sec. 5.6). Relighting the scene also brings out similar problems.

2 Three-Dimensional Perception and Visual Fatigue

In traditional 2-D cinema, since the result is always a 2D moving picture, almost anything can be filmed and displayed, without any effect on the spectator's health (except maybe for light flashes and stroboscopic effects), and the artist has a total freedom on what can be shown to the spectator. The result will appear as a moving picture drawn on a plane placed at some distance from the spectator, and will always be physically plausible.

In stereoscopic cinema, the two images need to be mutually consistent, so that a 3-D scene (real or virtual) can be reconstructed by the human brain. This implies strong geometric and photometric constraints between the images displayed to the left and the right eye, so that the 3-D shape of the scene is perceived correctly, and there is not too much strain on the human visual system which would result in visual fatigue.

Our goal in this chapter is to deal with the issues related to geometry and visual fatigue in stereoscopic cinema, but without any artistic considerations. We will just try to define bounds within which the artistic creativity can freely wander. These bounds were almost non-existent in 2-D movies, but as we will see they are crucial in stereoscopic movies.

Besides, since 3-D perception is naturally more important in stereoscopic movies, we have to understand what are the different visual features that produce depth perception. Those features will be called *depth cues*, and surprisingly most are monoscopic and can be experienced by viewing a 2-D image. Stereoscopy is only one depth cue amongst many others, although it may be the most complicated to deal with. The stereographer Phil Streather, quoting Lenny Lipton, said: "Good 3D is not just about setting a good background. You need to pay good attention to the seven monocular cues – aerial perspective, inter position, light and shade, relative size, texture gradients, perspective and motion parallax. Artists have used the first five of those cues for centuries. The final stage is depth balancing."

2.1 Monoscopic Depth Cues

The perception of 3-D shape is caused by the co-occurrence of a number of consistent visual artifacts called depth cues. These depth cues can be split into monoscopic cues, and stereoscopy (or stereopsis). For a review of 3-D shape perception from the cognitive science perspective, see Todd [65]. The basic seven monoscopic depth cues, as described in Lipton [37, chap. 2] (see also [61]), and illustrated Fig. 1, are:

- *Light and shade*
- *Relative size*, or retinal image size size (smaller objects are farther)
- *Interposition*, or overlapping (overlapped objects lie behind)
- *Textural gradient* (increase in density of a projected texture as a function of distance and slant)
- *Aerial perspective* (usually caused by haze)
- *Motion parallax* (2-D motion of closer objects is faster)
- *Perspective*, or linear perspective

As can bee seen in a famous drawing by Hogarth (Fig. 2), their importance can be easily demonstrated by using contradictory depth cues.

Lipton [37] also refers to what he calls a "physiological cue": *Accommodation* (the monoscopic focus response of the eye, or how much the ciliary muscles contract to maintain a clear image of an object as its distance changes). However, it is not clear from psychophysics experiments whether this should be considered as depth cue, i.e. if it gives an indication of depth in the absence of any other depth cue.

Although it is usually forgotten in the list of depth cues, we should also add *depth of field*, or retinal image blur [71, 28] (it is different from the accommodation

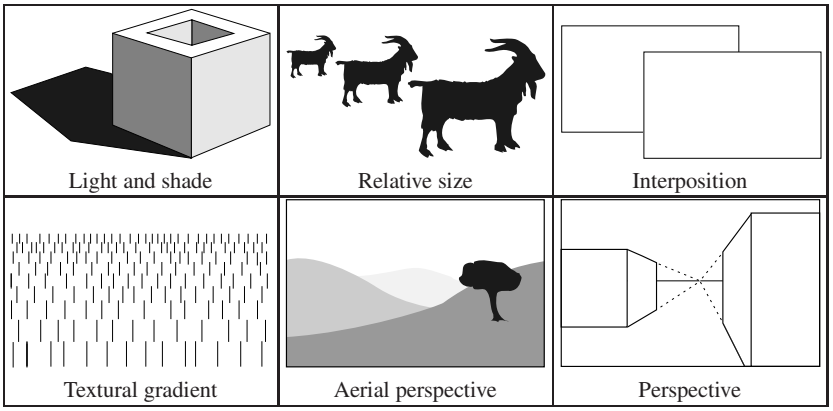


Fig. 1 Six monoscopic depth cues (from [61]). The seventh is motion parallax, which is hard to illustrate, and depth of field can also be considered as a depth cue (see Fig. 3).

cue cited before, which refers to the accommodation *distance* only, not to the deph of field), the importance of which is well illustrated by Fig.3. The depth of field of the Human eye is around 0.3 Diopters (D) in normal situations, although finer studies [40] claim that it also slightly depends on parameters such as the pupil size, wavelength, and spectral composition. Diopters are inverse of meters: at a focus distance of 3m, a depth of field of $\pm$ 0.3D, means that the in-focus range is from $1/(\frac{1}{3}+0.3) \approx 1.6$m to $1/(\frac{1}{3}-0.3) = 30$m, whereas at a focus distance of 30cm, the in-focus depth range is only from 27.5cm to 33cm (it is easy to understand from this formula why we prefer using diopters rather than a distance range to measure the depth of field: diopters are independent of the focus distance, and can easily be converted to a distance range). This explains why the photograph in Fig. 3 looks like a model rather than an actual-size scene [27]: the in-focus parts of the scene seem to be only about 30cm away from the spectator. The depth of field range is not much affected by age, so this depth cue may be learned from observations over a long period, whereas the accommodation range goes from 12D for children, to 8D for young adults, to... below 1D for presbyopes.

2.2 Stereoscopy and Stereopsis

Stereoscopy, i.e. the fact that we are looking at a scene using our two eyes, brings two additional physiological cues [37]:

- *Vergence* (the angle between the line-of-sight of both eyes);
- *Disparity* (the positional difference between the two retinal images of a scene point, which is non-zero for objects behind or in front of the convergence point).

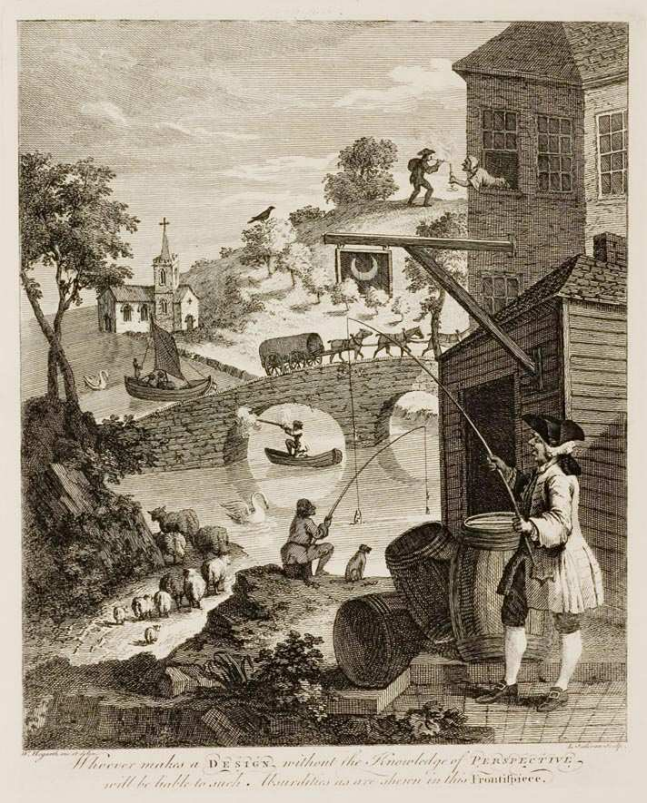


Fig. 2 "Whoever makes a DESIGN without the Knowledge of PERSPECTIVE will be liable to such Absurdities as are shown in this Frontispiece" (engraving by William Hogarth, 1754), a proof by contradiction of the importance of many monoscopic depth cues.

These cues are used by the perception process called stereopsis, which gives a sensation of depth from two different viewpoints, mainly from the horizontal disparity.

Although stereoscopy and motion parallax are very powerful 3D depth cues, it should be noted that human observers asked to make judgments about the 3-D metric structure of a scene from these cues are usually subject to large systematic errors [66, 63].

2.3 Conflicting Cues

All these cues (the 8 monoscopic cues and stereopsis) may be conflicting, i.e. giving opposite indications on the scene geometry. Many optical illusions make heavy use of these conflicting cues, i.e. when an object seems smaller because lines in the



Fig. 3 Focus matters! This photo is of a real scene, but the depth of field was reduced by a tilt-shift effect to make it look like a model [26] (photo by oseillo).

image suggest a vanishing point. Two famous examples are the Ames room and the pseudoscope.

The Ames room (invented by Adelbert Ames in 1934) is an example where monocular cues are conflicting. Ames room is contained in a large box, and the spectator can look at it though a single viewpoint, which is a hole in one of the walls. From this viewpoint, this room seems to be cubic-shaped because converging lines in the scene suggest the three standard lines directions (the vertical, and two orthogonal horizontal directions), but it is really trapezoidal (fig. 4). Perspective cues are influenced by prior knowledge of what a room should look like, so that persons standing in each far corner of the room will appear to be either very small or very big. The room itself can be seen from a peep hole at the front, thus forbidding binocular vision. In this precise case, binocular vision would easily disambiguate the conflicting cues by concluding that the room is not cubic. The Ames room was used in movies such as *Eternal Sunshine of the Spotless Minds* by Michel Gondry or the *Lord of the Rings* trilogy.

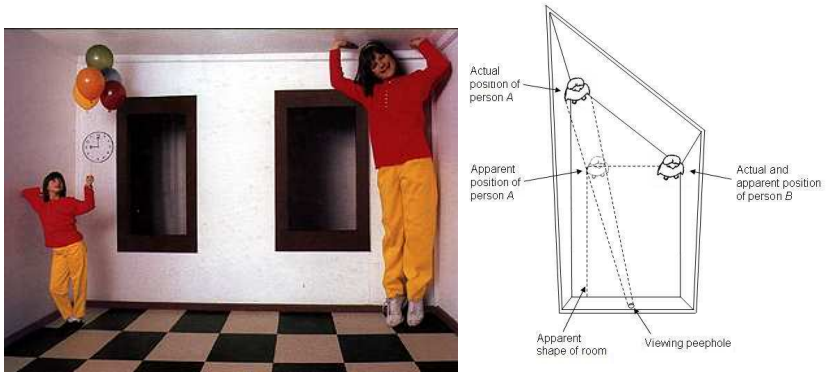


Fig. 4 Ames room: an example of conflicting perspective and relative size cues.

Another example of conflicting cues, which is more related to stereoscopic cinema, is illustrated by the pseudoscope. The pseudoscope (invented by Charles Wheatstone) is an binocular device which switches the viewpoints from the left and right eyes, so that all stereoscopic cues are reversed, but the monoscopic cues still remain and usually dominate the stereoscopic cues. The viewer still has the impression of "seeing in 3-D", and the closer objects in the scene actually seem bigger than they are, because the binocular disparity indicates that these big objects (in the image) are far away. This situation happened quite often during the projection of stereoscopic movies in the past [83], where the filters in front of the projectors or the film reels were accidentally reversed, but the audience usually did not notice what was wrong, and still had the impression of having seen a 3-D movie, though they thought it probably was a bad one because of the resulting headache.

Conflicting perspective and stereoscopic cues were actually used heavily by Pete Kozachik, Director of Photography on Henry Selick's stereoscopic film Coraline, to give the audience different sensations [34, 11]: "Henry wanted to create a sense of confinement to suggest Coraline's feelings of loneliness and boredom in her new home. His idea had interiors built with a strong forced perspective and shot in 3-D to give conflicting cues on how deep the rooms really were. Later, we see establishing shots of the more appealing Other World rooms shot from the same position but built with normal perspective. The compositions match in 2-D, but the 3-D depth cues evoke a different feel for each room."

2.4 Inconsistent Cues

Inconsistent cues are usually less disturbing for the spectator than conflicting cues. They are defined as cues that indicate different amounts of depth in the same direction. They have been used for ages in bas-relief, where the lighting cue enhances the depth perceived by the binocular system, and as a matter of fact bas-relief is

usually better appreciated from a far distance, where the stereoscopic cues have less importance.

An effect that is often observed when looking at stereoscopic photographs is called the *cardboard effect* [77, 41]: some depth is clearly perceived, but the amount of depth is too small with respect to the expected depth from the image size of the objects, resulting in objects appearing as flat, or drawn on cardboard of billboards. We will explain later how to predict this effect, and most importantly how to avoid it.

Another well-known stereoscopic effect is called the *puppet-theater effect* (also called *pinching*): background objects do not appear as small as expected, so that foreground objects appear proportionately smaller.

These inconsistent cues can easily be avoided if there is total control on the shooting geometry, including camera placement. If there are some unavoidable constraints on the shooting geometry, we will explain in Sec. 5 how *some* of these inconsistent cues related to stereoscopy can be corrected in post-production (Sec. 5).

2.5 Sources of Visual Fatigue

Visual fatigue is probably the most important point to be considered in stereoscopic cinema. Stereoscopic movies in the past often resulted in a bad viewing experience, and this reduced a lot the acceptance of stereoscopic cinema by a large public. Ukai and Howarth [67] produced a reference study on visual fatigue caused by viewing stereoscopic films, and is a good introduction to this field.

The symptoms of visual fatigue may be conscious (headache, tiredness, soreness of the eyes) or unconscious (perturbation of the oculo-motor system). It should actually be considered as a public health concern [69], just as the critical fusion frequency on CRT screens 50 years ago, as it may actually lead to difficulties in judging distances (which is very important in such tasks as driving). Ukai and Howarth [67, Sec. 6] even report the case of an infant whose oculo-motor system was permanently disturbed by viewing a stereoscopic movie. Although the long-term effects of viewing stereoscopic cinema were not studied due to the fact that this medium is not yet widespread, many studies exist on the effects on health of using virtual reality displays [70, 69, 35]. Virtual reality displays are widely used in the industry (from desktop displays to immersive displays), and are sometimes used daily by people working in industrial design, data visualization, or simulation.

The sources of visual fatigue that are specific to stereoscopic motion pictures are mainly due to binocular asymmetry, i.e. photometric or geometric differences between the left and right retinal images. Kooi and Toet [33] experimentally measured thresholds on the various assymetries that will lead to visual incomfort (incomfort is the lowest grade of conscious visual fatigue). For example, they measured, in agreement with Pastoor's rule of thumb [47] that a 35 arcmin horizontal disparity range is quite acceptable for binocular perception of 3-D and 70 arcmin disparity is too much to be viewed. They also found out that the human visual system is most sen-

sitive to vertical binocular disparities. The various quantitative binocular thresholds computed from their experiments can be found in Kooi and Toet [33, Table 4].

The main sources of visual fatigue can be listed as:

- *Crosstalk* (sometimes called *crossover* or *ghosting*), which is usually due to a stereoscopic viewing system with a single screen: a small fraction of the intensity from the left image can be seen in the right eye, and vice-versa. The typical values for crosstalk [33] are 0.1-0.3% with polarization-based systems, and 4-10% with LCD shutter glasses. Preprocessing can be applied to the images before displaying in order to reduce crosstalk by subtracting a fraction of the left image from the right image [42] – a process sometimes called *ghost-busting*.
- *Breaking the proscenium rule* (or breaking the stereoscopic window) happens when there are interposition errors between the stereoscopic imagery and the edges of the display (see Fig. 5 and Mendiburu [42, Chap. 5]). A simple way to correct this is to move the proscenium closer to the spectator by adding borders to the image, but this is not always as easy as it seems (see Sec. 5.5).
- *Horizontal disparity limits* are the minimum and maximum disparity values that can be accepted without producing visual fatigue. An obvious bound for disparity is that the eyes should not diverge. Another limit concerns the range of acceptable disparities within a stereoscopic scene that can be fused simultaneously by the human visual system.
- *Vertical disparity* causes torsion motion of the ocular globes, and is only tolerable for short time intervals. Our oculomotor system learned the epipolar geometry of our eyes over a lifetime of real-world experience, and any deviation from the learned motion causes strain.
- *Vergence-accomodation conflicts* occur when the focus distance of the eyes is not consistent with their vergence angle. It happens quite often when viewing stereoscopic cinema, since the display usually consists of a planar surface placed at a fixed distance. Strictly speaking, any 3-D point that is not in the convergence plane will have an accomodation distance, which is exactly the screen distance, different from the vergence distance. However, this constraint can be somewhat relaxed by using the depth of field of the visual system, as will be seen later.

Geometric asymmetries come very often either from a misalignment or from a difference between the optics in the camera system or in the projection system, as seen in Fig. 6. In the following, we will only discuss *horizontal disparity limits*, *vertical disparity*, and *vergence-accommodation conflicts*, since the other sources of visual fatigue are easier to deal with.

2.5.1 Horizontal Disparity Limits

The most simple and obvious disparity limit is eye divergence. In their early work on stereoscopic cinema, the Spottiswoodes said: “It is found that divergence is likely to cause eyestrain, and therefore screen parallaxes in excess of the eye separation

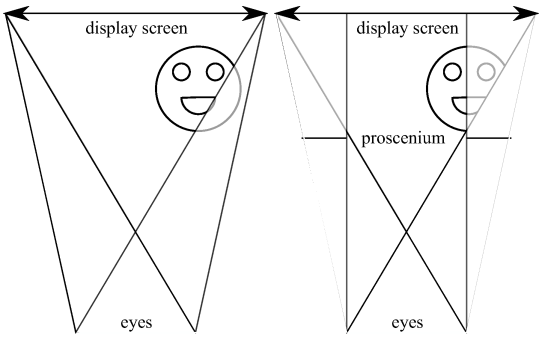


Fig. 5 Breaking the proscenium rule: (left) part of the object in front of the proscenium arch is not visible in one eye, which breaks the proscenium rule; (right) masking part of the image in each eye moves the proscenium closer than the object, and the proscenium rule is re-established.

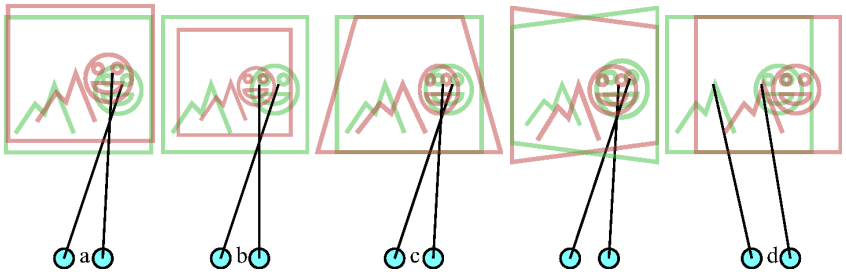


Fig. 6 A few examples of geometric asymmetries: (a) Vertical shift, (b) Size or magnification difference, (c) Distortion difference, (d) Keystone distortion due to toed-in cameras, (e) Horizontal shift - leading to eye divergence in this case (adapted from Ukai and Howarth [67]).

should be avoided". But they also went on to say, in listing future development requirements, that "Much experimental work must be carried out to determine limiting values of divergence at different viewing distances which are acceptable without eyestrain".

These limiting values are the maximum disparities acceptable around the convergence point, usually expressed as angular values, such that the binocular fusion of the 3-D scene is performed without any form of eyestrain or visual fatigue. Many publications dealt with the subject of finding the horizontal disparity limits [79, 30, 47].

The horizontal disparity limits are actually closely related to the depth of field, as noted by Lambooij et al [35]: "An accepted limit for DOF in optical power for a 3 mm pupil diameter (common under normal daylight conditions) and the eyes focusing at infinity, is one-third of a diopter. With respect to the revisited Panum's fusion area[3], disparities beyond one degree (a conservative application of the 60 to 70 arcmin recommendation), are assumed to cause visual discomfort, which actu-

[3] In the human visual system, the space around the current fixation point which can be fused is called Panum's area or fusion area. It is usually measured in minutes of arc (arcmin).

ally results from the human eye's aperture and depth of field. Though this nowadays serves as a rule-of-thumb, it is acknowledged as a limit, because lower recommendations have been reported as well. If both the limits of disparity and DOF are calculated in distances, they show very high resemblance.". Yano et al [78] also showed that images containing disparities beyond the depth of field ($\pm$ 0.2D depth of field, which means $\pm 0.82^\circ$ in disparity) cause visual fatigue.

2.5.2 Vertical Disparity

Let us suppose that the line joining both eyes is horizontal, and that the stereoscopic display screen is vertical and parallel to this line. The images of any 3-D point projected onto the display screen using each eye optical center as the centers of projection are two points which are aligned horizontally, i.e. have no vertical disparity. Thus all the scene points that are displayed on the screen should have no vertical disparity. Vertical disparity (see Fig. 6 for some examples) may come from a misalignment of the cameras or of the display devices, from a focal length difference between the optics of the cameras, from keystone distortion, due to a toed-in camera configuration, or from nonlinear (e.g. radial) distortions.

However, it is to be noted that vertical disparities exist in the visual system: remember that the eye is not a linear perspective camera, but a spheric sensor, so that an object which is not in the median plane between both eyes will be closer to one eye than to the other, and thus its image will be bigger in one eye than in the other (a spheric sensor basically measures angles). The size ratio between the two images is called the vertical size ratio, or VSR. VSR is naturally present when rectified images (i.e. with no vertical disparity) are projected on a flat display screen: a vertical rod, though it's displayed with the same size on the left and right images on the flat display, subtends a larger angle in the nearest eye.

Psychophysical experiments showed that vertical disparity gradients have a strong influence on the perception of stereoscopic shape, depth and size [44, 49, 5]. For example, the so-called induced-size effect [44] is caused by a vertical gradient of vertical disparity (vertical-size parallax transformation) between the half images of an isolated surface, which creates an impression of a surface slanted in depth. Ogle [44, 45] called it the induced-size effect because it is as though the vertical magnification of the image in one eye induces an equivalent horizontal magnification of the image in the other eye[4].

Allison [3] also notes that vertical disparities can be used to fool the visual system: "Images on the retinae of a fronto-parallel plane placed at some distance actually have some keystone distortion, which may be used as a depth cue. Displaying keystone-distorted images on that fronto-parallel screen actually exaggerates that keystone distortion when the viewer is focusing on the center of the screen, and would thus giving a cue that the surface is nearer than the physical screen.".

[4] However, Read and Cumming [48] show that non-zero vertical disparity sensors are in fact not necessary to explain the induced size effect.

By displaying keystoned images, the VSR is distorted in a complicated way which may be inconsistent with the horizontal disparities. Besides, that distortion depends on the viewer position with respect to the screen. When the images displayed on the screen are rectified, although depth perception may be distorted depending on the viewer position, horizontal and vertical disparities will always be consistent, as long as the viewer's interocular (the line joining the two optical centers) is kept parallel to the screen and horizontal (this viewing position may be hard to obtain in the side rows of wide movie theaters).

Woods et al [73] discuss sources of distortion in stereo camera arrangements as well as the human factors considerations required when creating stereo images. These experiments show that there is a limit in the screen disparity which it is comfortable to show on stereoscopic displays. A limit of 10 mm screen disparity on a 16" display at a viewing distance of 800 mm was found to be the maximum that all 10 subjects of the experiment could view. Their main recommendation is to use a parallel camera configuration in preference to converged cameras, in order to avoid keystone distortion and depth plane curvature.

Is Vertical Disparity Really a Source of Visual Fatigue? From the fact that vertical disparities are actually a depth cue, there has been a debate on whether vertical disparities can be a source of visual fatigue, since they are naturally present in retinal images. Stelmach et al [60] and [56] claim that keystone and depth plane curvature cause minimal discomfort: images plane shift (equivalent to rectified images) and toed-in cameras are equally comfortable in their opinion. However, we must distinguish between visual discomfort, which is conscious, and visual fatigue, where the viewer may not be conscious of the problem during the experiment, but headache, eyestrain, or long-term effects can happen.

Even in the case where the vertical disparities are not due to a uniform transform of the images, such as a rotation, scaling or homography, Stevenson and Schor [62] demonstrated that human stereo matching does not actually follow the epipolar lines, and human subjects can still make accurate near/far depth discrimination when the vertical disparity range is as high as 45 arcmin.

Allison [4] concludes that, although keystone-distorted images coming from a toed-in camera configuration can be displayed with their vertical disparities without discomfort, the images should preferably be rectified, because the additional depth cues caused by keystone distortion perturb the actual depth perception process.

2.5.3 Vergence-Accommodation Conflicts

When looking at a real 3-D scene, the distance of accomodation (i.e. the focus distance) is equal to the distance of convergence, which is the distance to the perceived object. This relation between these two oculo-motor functions, called Donder's line [78], is learned through the first years of life, and is used by the visual system to quickly focus-and-converge on objects surrounding us. The relation between vergence and accomodation does not have to follow exactly Donder's line:

there is an area around it where vergence and accomodation agree, which is called Percival's zone of comfort [28, 78].

When viewing a stereoscopic movie, the distance of accomodation differs from the distance of convergence, which is the distance to the perceived object (Fig. 7). This discrepancy causes a perturbation of the oculo-motor system [78, 17], which causes visual fatigue, and may even damage the visual acuity, which is reported by Ukai and Howarth [67] to have plasticity until the age of 8 or later. This problem has been largely studied for virtual reality (VR) displays, and lately for 3-D television (3DTV), but has been largely overlooked in movie theater situations. In fact, many stereoscopic movies, especially IMAX-3D movies, make heavy use of spectacular effects by presenting perceived objects which are very close to the spectator, when the screen distance is about 20m.

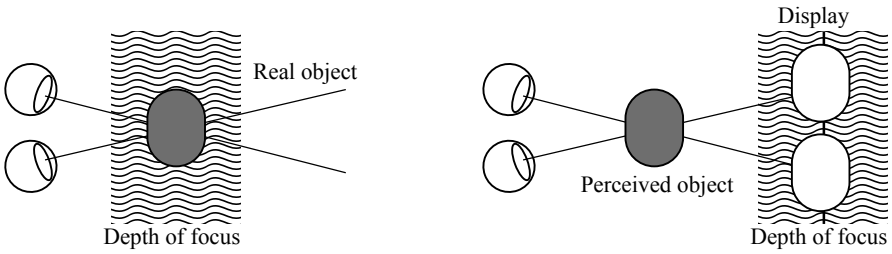


Fig. 7 Vergence and accomodation: they are consistent when viewing a real object (left), but may be conflicting when viewing a stereoscopic display, since the perceived object may not lie within the depth of field range (right). Adapted from Emoto et al [17].

Wann and Mon-Williams [69] cite it as one of the main sources of stress in VR displays. They observed that, in a situation where other sources of visual discomfort were eliminated, prolonged use of a stereoscopic display caused short-term modifications in the normal accommodation-vergence relationship.

Hoffman et al [28] designed a special 3-D display where vergence and focus cues are consistent [2]. The display is designed so that its depth of field approximately corresponds to the Human depth of field. By using this display, they showed that "when focus cues are correct or nearly correct, (1) the time required to identify a stereoscopic stimulus is reduced, (2) stereo-acuity in a time-limited task is increased, (3) distortions in perceived depth are reduced, and (4) viewer fatigue and discomfort are reduced." However, this screen is merely experimental, and it is impractical for home or movie theater use.

The depth of field may be converted to disparities in degrees using simple geometric reasoning. The depth of field of the human visual system is, depending on the authors, between $\pm$ 0.2D or $\pm 0.82^\circ$ [78] and $\pm$ 0.3D [12]. The limit to binocular fusion is from 2 to 3 degrees at the front or back of the stereoscopic display, and Percival's zone of comfort is about one third of this, i.e. 0.67 to 1 degree. We note that the in-focus range almost corresponds to Percival's zone of comfort for binoc-

ular fusion, which probably comes from the fact that the visual system only learned to fuse non-blurred objects within the in-focus range.

Let us take for example a conservative value of $\pm$ 0.2D for the depth of field. For a movie theater screen placed at 16m, the in-focus range is from $1/(1/16+0.2) \approx$ 3.8m to infinity, whereas for a 3DTV screen placed at 3.5m, it is from $1/(1/3.5+$ $0.2) \approx$ 2m to $1/(1/3.5+0.2) \approx 11.7$m. As we will see later, this means that the camera focus range should theoretically be different when shooting movies for a movie theater or a 3-D television.

3 Picking the Right Shooting Geometry

3.1 The Spottiswoode Point of View

Spottiswoode et al [57] wrote the first essay on the perceived geometry in stereoscopic cinema. They devised how depth is distorted by "stereoscopic transmission" (i.e. recording and reproduction of a stereoscopic movie), and how to achieve "continuity in space", or making sure that there is a smooth transition in depth when switching from one stereoscopic shot to another.

They did a strong criticism of the "human vision" systems (i.e. systems that were trying to mimic the human eyes, with a 6.5cm interocular, and a 0.3° convergence). They claimed that all stereoscopic parameters can be and should be adapted, either at shooting time or as post-corrections, depending on screen size, to get the desirable effects (depth magnification or reduction, and continuity in space). According to them, the main stereoscopic parameter is the *nearness factor* N, defined as the ratio between the viewing distance from screen and the distance to the fused image (with our notations, $N = H'/Z'$): "for any pair of optical image points, the ratio of the spectator's viewing distance from the screen to his distance to the fused image point is a constant, no matter whereabouts in the theater he may be sitting". Continuity in space is achieved by slowly shifting over a few seconds the images in the horizontal direction before and after the cuts. The spectator does not notice that the images are slowly shifting: the vergence angle of the eyes is adapted by the human visual system, and depth perception is almost the same. Due to the technical tools available at this time, this is the only post-correction method they propose.

They list three classes of stereoscopic transmission:

- *ortho-infinite*, where infinity points are correctly represented at infinity,
- *hyper-infinite*, where objects short of infinity are represented at infinity (which means that infinity points cause divergence),
- *hypo-infinite*, where objects at infinity are represented closer than infinity (which causes the cardboard effect, see Sec. 2.4).

In order to help the stereographer picking up the right shooting parameters, they invented a calculator, called the *Stereomeasure*, which computes the relation between the various parameters of 3-D recording and reproduction. They claim that a

stereoscopic movie has to be made for given projection conditions (screen size and distance to screen), which is a fact sometimes overlooked either by stereographers, or by the 3DTV industry (however, as discussed in Sec. 5, the stereoscopic movie may be adapted to other screen sizes and distances).

In the Spottiswoode setup, the standard distance from spectator to screen should be from $2W'$ to $2.5W'$ (W' is the screen width). They place the proscenium arch at $N = 2$ (half distance from screen), and almost everything in the scene happens between $N = 0$ (infinity) and $N = 2$ (half distance) ($N = 1$ is the screen plane).

Their work has been strongly criticized by many stereographers, sometimes with wrong arguments [37, 54], but the main problem is probably that the artists do not want to be constrained by mathematics when they are creating: cinematography has always been an art of freedom, with a few rules of thumb that could always be ignored. But the reality is here: the constraints on stereoscopic cinema are much stronger than on 2-D cinema, and bypassing the rules results in bad movies causing eyestrain or headache to the spectator. A bad stereoscopic movie can be a very good 2-D movie, but adding the stereoscopic dimension will always modify the perceived quality of the movie, either by adding a feeling of "being there", or by obfuscating the intrinsic qualities of the movie with ill-managed stereoscopy.

There are some problems with the Spottiswoode's theorisation of stereoscopic transmission, though, and this section will try to shed some light on some of these:

- the parametrization by the nearness factor hides the fact that strong nonlinear depth and size distortions may occur in some cases, especially on far points;
- divergence at infinity will happen quite often when images are shifted in order to achieve continuity in space;
- shifting the images may break the vergence-accomodation constraints, in particular Percival's zone of comfort, and cause visual fatigue.

Woods et al [73] extended this work and also computed spatial distortion of the perceived geometry. Although their study is more focused on determining horizontal and vertical disparity limits, they also studied depth plane curvature effects, where a fronto-parallel plane appears to be curved. This situation arises when non-rectified images are used and the camera configuration is toed-in (i.e. with a non-zero vergence angle).

More recently, Masaoka et al [41] from the NHK labs also did a similar study, and presented a software tool which is able to predict spatial distortions that happen when using given shooting and viewing parameters.

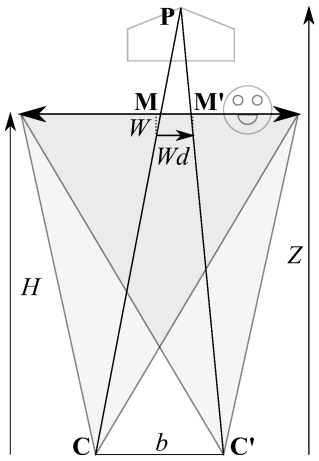
3.2 Shooting and Viewing Geometries

As was shown by Spottiswoode et al [57] in the early days of stereoscopic cinema, projecting a stereoscopic movie on different screen sizes and distances will produce different perceptions of depth.

One obvious solution was adopted by the large format stereoscopic cinema (IMAX 3-D): if the film is shot with parallel cameras (i.e. vergence is zero), and is projected with parallel projectors that have a human-like interocular (usually 6cm for IMAX-3D), then infinity points will always be perceived exactly at infinity, and divergence will never occur [83]. Large-format stereoscopic cinema has less constraints than stereoscopic cinema targeted at standard movie theaters, since the screen is practically borderless, and the audience is located near the center of the hemispherical screen. The camera interocular is usually close to the human interocular, but it may be played with easily, depending on the scene to be shot (as in *Bugs! 3-D* by Phil Streather, where hypostereo or gigantism is heavily used). If the camera interocular is the same as the human interocular, then everything in the scene will appear at the same depth and size as if seen with normal vision, which is a very pleasant experience.

However, shooting parallel is not always advisable or even possible for standard stereoscopic movies. Close scenes for example, where the subject is closer than the movie theater screen, require camera convergence, or the film subject will appear too close to the spectator and will break the proscenium rule most of the time. But using camera convergence has many disadvantages, and we will see that it may cause heavy distortions on the 3-D scene.

Let us study the distortions caused by given shooting and viewing geometries. The geometric parameters we use (Fig. 8) are very simple but describe fully the stereoscopic setup. Compared to camera-based parameters, as used by Yamanoue et al [77], we can easily attach a simple meaning to each parameter and understand its effect on space distortions. We assume that the stereoscopic movie is rectified and thus contains no vertical disparity (see Sec. 5.1), 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
C, **C**′	camera optical center	eye optical center
P	physical scene point	perceived 3-D point
M, **M**′	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	disparity (as a fraction of W)	

Fig. 8 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 fronto-parallel (or width and height) directions, and in the depth direction. Spottiswoode et al [57] defined the *shape ratio* as the ratio between depth magnification and width magnification, Yamanoue et al [77] call it E_p or *depth reduction*, and Mendiburu [42] uses the term *roundness factor*. We will use *roundness factor* in the remaining of our study. A low roundness factor will result in 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. 8. Let d be the disparity in the images. The disparity on the display screen is $d' = d + d_0$, taking into account an optional shift d_0 between the images (shifting can be done at the shooting stage, in post-production, or at the display stage).

Triangles $\mathbf{MPM'}$ and $\mathbf{CPC'}$ are homothetic, consequently:

$$\frac{Z-H}{Z} = \frac{W}{b}d. \tag{1}$$

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

$$d = \frac{b}{W}\frac{Z-H}{Z}, \text{ or } Z = \frac{H}{1-\frac{W}{b}d}, \tag{2}$$

and the perceived depth Z' as a function of the disparity d':

$$Z' = \frac{H'}{1-\frac{W'}{b'}(d+d_0)}. \tag{3}$$

Finally, eliminating the disparity d from both equations gives the relation between real depth Z and perceived depth Z':

$$Z' = \frac{H'}{1-\frac{W'}{b'}(\frac{b}{W}\frac{Z-H}{Z}+d_0)} \quad \text{or} \quad Z = \frac{H}{1-\frac{W}{b}(\frac{b'}{W'}\frac{Z'-H'}{Z'}-d_0)} \tag{4}$$

3.3 Depth Distortions

Let us now compute the depth distortion from the perceived depth. In the general case, points at infinity ($Z \to +\infty$) are perceived at:

$$Z' = \frac{H'}{1-\frac{W'}{b'}(\frac{b}{W}+d_0)}. \tag{5}$$

Eye divergence happens when Z' becomes negative. The eyes diverge when looking at scene points at infinity ($Z \to +\infty$) if and only if:

$$\frac{b'}{W'} < \frac{b}{W} + d_0. \tag{6}$$

In this case, the real depth which is mapped to $Z' = \infty$ can be computed from (4) as:

$$Z = \frac{H}{1 - \frac{W}{b}\left(\frac{b'}{W'} - d_0\right)}, \tag{7}$$

and any object at a depth beyond this one will cause divergence.

The relation between Z and Z' is nonlinear, except if $\frac{W}{b} = \frac{W'}{b'}$ and $d_0 = 0$, which we call the *canonical setup*. In this case, the relation between Z and Z' simplifies to:

$$Z' = Z\frac{H'}{H}. \qquad \text{(Canonical setup: } \frac{W}{b} = \frac{W'}{b'}, d_0 = 0) \tag{8}$$

Let us now study how depth distortion behaves when we are not in the canonical setup anymore. For our case study, we start from a purely canonical setup, with $b = b' = 6.5cm$, $W = W' = 10m$, $H = H' = 15m$ and $d_0 = 0$. In the following charts, depth is measured from the convergence plane or the screen plane (depth increases away from the cameras/eyes).

The first chart (Fig. 9) shows the effect of changing only the camera interocular b (the distance to the convergence plane H remains unchanged, which implies that the vergence angle changes accordingly). This shows the well-known effects called *hyperstereo* and *hypostereo*: the roundness factor of in-screen object varies from high values (hyperstereo) to low values (hypostereo).

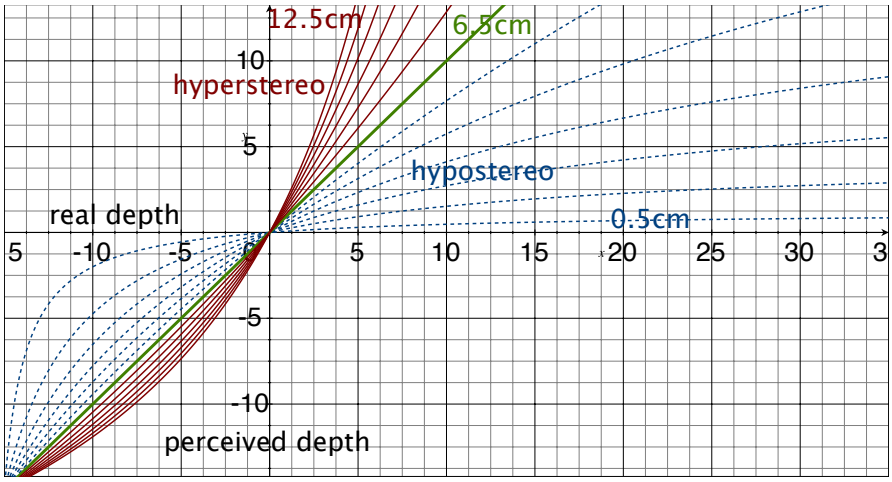


Fig. 9 Perceived depth as a function of real depth for different values of the camera interocular b. This graph demonstrates the well-known hyperstereo and hypostereo phenomenon.

Let us now suppose that the subject being filmed is moving away from the camera, and we want to keep its image size constant by adjusting the zoom and vergence accordingly, i.e. W remains constant. To keep the object's roundness factor constant, we also keep the interocular b proportional to the object distance H ($b = \alpha H$). The effect on perceived depth is shown in Fig. 10. We see that the depth magnification or roundness factor close to the screen plane remains equal to 1, as expected, but the depth of out-of-screen objects is distorted, and a closer analysis would show that the wide-interocular-zoomed-in configuration creates a *puppet-theater effect*, since farther objects have a larger image size than expected. This is one configuration where the roundness factor of the in-screen objects can be kept constant while changing the stereoscopic camera parameters, and in the next section we will show how to compute all the changes in camera parameters that give a similar result.

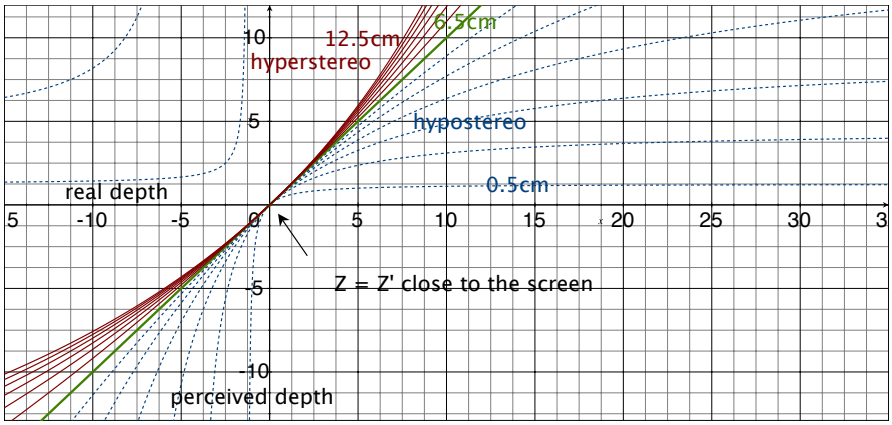


Fig. 10 When the object moves away, but we keep the object image size constant by zooming in (i.e. H varies but W is constant), if we keep the camera baseline b proportional to convergence distance H, the depth magnification close to the screen plane remains equal to 1. But be careful: divergence may happen!

3.4 Shape Distortions and the Depth Consistency Rule

If we want the camera setup and the display setup to preserve the shape of all observed objects up to a global 3-D scale factor, a first constraint is that there must be a linear relation between depths, so we must be in the canonical setup described by eq. (8). A second constraint is that the ratio between depth magnification and image magnification, called the *roundness factor*, must be equal to 1. This means that the *only* configuration with faithful depth reproduction is when the shooting and viewing geometries are *homothetic* (i.e. there is a scale factor between the two

geometries):

$$\frac{W'}{W} = \frac{H'}{H} = \frac{b'}{b}. \qquad \text{(Homothetic configuration)} \tag{9}$$

In general, the depth and space distortion is nonlinear: it can easily be shown that the perceived space is a homographic transform of the real space. Shooting from farther away while zooming in with a bigger interocular doesn't distort (much) depth, as we showed in the previous section, and that's probably the right way to zoom in - the baseline should be proportional to the convergence distance. But one has to take care about the fact that the infinity plane in scene space may cause divergence.

More generally, we can introduce the *depth consistency rule*: Objects which are close to the convergence plane in the real scene or close to the screen in projection space ($Z = H$ or $Z' = H'$) should have a depth which is consistent with their apparent size, i.e. their roundness factor should be equal to 1.

The depth ratio between scene space and projection space close to the convergence plane can be computed from (4) as:

$$\frac{\partial Z'}{\partial Z}(Z = H) = \frac{b}{HW}\frac{H'W'}{b'}, \tag{10}$$

and the apparent size ratio is simply $\frac{W'}{W}$. Setting the ratio between both, i.e. the roundness factor, equal to 1 leads to the *depth consistency rule*:

$$\frac{b}{H} = \frac{b'}{H'}. \qquad \text{(Depth consistency rule)} \tag{11}$$

One important fact arises from the depth consistency rule: for objects that are close to the convergence plane or the screen plane, screen size (W') does not matter, whereas most spectators expect to get exaggerated 3-D effects when looking at a movie on a bigger screen. Be careful, though, that for objects farther than the screen, especially at $Z \to +\infty$, divergence may occur on bigger screens. Since neither b, H, or b' can be changed when projecting a 3-D movie, the key parameter for the depth consistency rule is in fact screen distance, which dramatically influences the perceived depth: the bigger the screen distance, the higher the roundness factor will be. This means that if a stereoscopic movie, which was made to be seen in a movie theater from a distance of 16m, is viewed on a 3DTV from a distance of 4m, the roundness factor will be divided by 4 and the spectator will experience the classical *cardboard effect* where objects look flat.

What can we do to enforce the depth consistency rule, i.e. to produce the same 3-D experience in different environments, given the fact that b is fixed and H' is usually constrained by the viewing conditions (movie theater vs. home cinema vs. TV)? The only possible solution consists in artificially changing the shooting parameters in post production using techniques that will be described in Sec. 5.

3.5 Shooting With the Right Depth of Field

When disparities outside of the Percival zone of comfort are present, Ukai and Howarth [67] showed that vergence-accomodation conflicts that arise can be attenuated by reducing the depth of field: the ideal focus distance should be on the plane of convergence (i.e. the screen), and the depth of field should match the expected depth of field of the viewing conditions. They even showed that objects that are out of this focus range are surprisingly better perceived in 3-D if they are blurred than if they are in focus.

We saw (Sec. 2.1) that the human eye depth of field in normal conditions is between 0.2D and 0.3D (diopters) – let us say 0.2D to be conservative. This depth of field should be converted to a depth range in the targeted viewing conditions: $Z'_{\min} = 1/\left(\frac{1}{H'} - 0.2\right)$, $Z'_{\max} = 1/\left(\frac{1}{H} + 0.2\right)$ (if $Z'_{\max}$ is negative, then it is considered to be at infinity). Then, the perceived depth range should be converted to real depth range $[Z_{min}, Z_{max}]$, using the inverse of formula (4), and the camera aperture should be computed from this depth range.

Unfortunately, the viewing conditions are usually not known at shooting time, or the movie has to be made for a various range of screen distances and sizes. We will describe a possible solution to this general case in Sec. 5.4.

3.6 Remaining Issues

Even when the targeted viewing conditions are known precisely, the shooting conditions are sometimes constrained: for example, when filming wild animals, a minimum distance may be necessary, resulting in using a wide baseline and a long focal length (i.e. a large $\frac{b}{W}$). The resulting stereoscopic movie, although it will have a correct roundness factor around the screen plane, will probably have strong divergence at infinity, since $\frac{b'}{W'} < \frac{b}{W}$, which may look strange, even if the right depth of field is used.

Another kind of problem may happen, even with objects which are close to the screen plane: Psychophysics experiments showed that specular reflections may also be used as a shape cue, and more specifically as a curvature cue [8, 9]. Except when using a purely homothetic configuration, these depths cues will be inconsistent with other cues, even near the screen plane, and this effect will be particularly visible when using long focal lengths. This problem can probably not be overcome by changing the shooting geometry, and the best solution is probably to edit manually the specular reflections in post-production to make them look more natural, or to use the right makeup on the actors...

Inconsistent specular reflections can also be due to the use of a mirror rig (a stereoscopic camera rig using a half-silvered mirror to separate images for the left and right cameras): specular reflections are usually polarized by nature, and the transmission and reflection coefficients of the mirror depend on the polarization of

incoming light. As a result, specular reflections have a different aspect in the left and right images.

4 Lessons for Live-Action Stereoscopic Cinema from Animated 3-D

Shooting live-action stereoscopic cinema has progressed over the last decade with the arrival of more versatile camera rigs. Older stereo rigs had two fixed cameras, and the stereo extrinsic parameters – baseline and vergence – were changed by manual adjustment between shots. Newer rigs allow dynamic change of the stereo parameters during a shot. This still falls short of a final stage of versatility, which is the modification of the left- and right-eye images during post-production, to effectively change the stereo extrinsic parameters – in other words, this ultimate goal is to use the shot footage as a basis for synthesizing left- and right-eye images with any required stereo baseline and vergence.

New technologies are bringing us closer to the goal of synthesizing left- and right-eye views with different stereo parameters during post-production. But what are the benefits, and how will this create a better viewer experience? To answer this question, in this section we look at current practices in animated stereoscopic cinema. In animation, the creative team has complete control over camera position, camera motion, camera intrinsics, stereo extrinsics, and the 3-D structure of the scene. This allows scope to experiment with the stereoscopic experience, including doing 3-D manipulations and distortions that do not correspond to any physical reality but which enhance viewer experience. We describe four core techniques of animated stereoscopic cinema. The challenge for live-action 3-D is to create new technologies so that these techniques can be applied to live-action footage as easily as they are currently applied to animation.

4.1 Proscenium Arch or Floating Window

The proscenium arch or floating window was introduced earlier. The simplest way to project stereoscopic imagery is to capture images from a left- and right- camera and then put those images directly onto the cinema screen. This imposes a specific epipolar geometry on the viewer, and our eyes adjust to it. Now consider the four-sided boundary of the physical cinema screen. It also imposes epipolar constraints on the eyes (in fact these are epipolar constraints which are consistent with the whole rest of the physical world). But note that there is no reason why the epipolar geometry imposed by the stereoscopic images will be consistent with the epipolar constraints associated with the screen boundary. The result is conflicting visual cues.

The solution is to black-mask the two stereoscopic images. Consider the basic 3-D animation setup, and two cameras observing a 3-D scene. Now imagine be-

tween the camera and the scene a rectangular window, in 3-D, so that the cameras view the scene through the window, but the window is surrounded by a black wall where nothing is visible. This is the proscenium arch or floating window, a virtual 3-D entity interjected between the cameras and the scene. The visual effect of the window is achieved by black-masking the boundaries of the left- and right- eye images. Since this floating window is a 3-D entity that is consistent with the cameras and the rest of the 3-D scene, it does not give rise to conflicting visual cues.

4.2 Floating Windows and Audience Experience

Section 4.1 described a basic motivation for the floating window, motivated by comfort in the viewing experience. There is a further use for the technique. Note that the window can be placed anywhere in the 3-D scene, It can lie between the camera and the scene, part-way through the scene, or behind the scene. This is not perceived explicitly but placing the 3-D scene behind the floating window relative to the audience produces a more subdued passive feeling, according to accepted opinion in the creative community. Placing the scene in front of the floating window relative to the audience produces a more engaged active feeling. Also note that the window does not need to be fronto-parallel to the viewer. Instead it can be tilted in 3-D space. Again the viewer is typically unaware of this consciously, but orientation of the floating window can produce subliminal effects e.g. a forward tilt of the upper part of the window can produce a looming feeling.

4.3 Window Violations

The topic of floating windows leads naturally to another technique. Consider again the 3-D setup - cameras, floating window, and 3-D scene. First consider objects that are on the far side of the window from the cameras. When they are center-stage, they are visible through the window. As they move off-stage and are obscured by the surrounding wall of the virtual window. This is all consistent from the viewpoint of the two stereoscopic cameras that are viewing the scene. Now consider an object that lies between the cameras and the window. While it's center-stage and visible in the two eyes, everything is fine. As it moves off-stage, it intersects the view-frustrums created by each camera and the floating window, and nothing is visible outside the frustrum. But this is inconsistent to the eye – it's as though we are looking at someone in front of a doorway and their silhouette disappears as it passes over the view frustrums in the left- and right-eyes of the doorway which is to the rear of them.

The solution to this problem is simply not to allow such violations. This is straightforward in animation but in live-action stereoscopic cinema, of course, it requires the non-trivial recovery of the stereo camera positions and the 3-D scene to detect when this is happening, and avoid it.

4.4 Multi-Rigging

So far, we considered manipulations of the left- and right- eye images that are consistent with a physical 3-D reality. Multi-rigging, however, is a technique for creating stereoscopic images which produce a desired viewing experience, but could not have arisen from a physically correct 3-D situation. Consider a scene with objects A and B. The left camera is kept fixed but there are two right cameras, one for shooting object A and one for shooting object B. Two different right eye images are generated, and they are then composited so that objects A and B appear in a single composite right eye image.

Why do this? Consider the case where A is close to the camera and B is far away. In a normal setup, B will appear flat due to distance. But by using a large stereo baseline for shooting object B, it is possible to capture more information around the occluding contour of the object, and give it a greater feeling of roundedness, even though it is placed more distantly in the scene. Again this is straightforward for animated stereoscopic cinema, but requires 3-D capture of the scene, and the ability to modify stereo baseline at post-production time, to apply it to live-action stereoscopic cinema.

5 Post-production of Stereoscopic Movies

When filming with stereoscopic cameras, even if the left and right views are rectified in post-production (Sec. 5.1), there are many reasons for which the movie may not be adapted to given viewing conditions, among which:

- The screen distance and screen size are different from the ones the movie was filmed for, resulting in a different *roundness factor* for objects close to the screen.
- Because of a screen size larger than expected, the points at infinity cause eye divergence.
- There were constraints on camera placement (as those that happen when filming sports or wildlife), which cause large disparities, or even divergence, on far-away objects.
- The stereoscopic camera was not adjusted properly when filming.

As we will see during this section, changing the shooting geometry in post-production is theoretically possible, although the advanced techniques require high-quality computer vision and computer graphics algorithms to perform a process called view interpolation. Due to the fact that 3-D information can be extracted from the stereoscopic movies, there are also a few other stereo-specific post-production processes that may be improved by using computer vision techniques.

5.1 Eliminating Vertical Disparity: Rectification of Stereoscopic Movies

Vertical disparity is one of the sources of visual fatigue in stereoscopic cinema, and it may come from misaligned cameras or optics, or from toed-in camera configurations, where vertical disparity cannot be avoided.

In Computer Vision, 3-D reconstruction from stereoscopic images is usually preceded by a transformation of the original images [19, Chap. 12]. This transformation, called *rectification* is a 2-D warp of the images that aligns matching points on the same y coordinate in the two warped images. The rectification of a stereoscopic pair of images is usually preceded by the computation of the *epipolar geometry* of the stereoscopic camera system. Knowing the epipolar geometry, one can map a point in one of the images, to a line or curve in the other image which is the projection of the optical ray issued from that point onto that image. The rectification process transforms the epipolar lines or curves into horizontal lines, so that stereoscopic matching by computer vision algorithms is made easier.

It appears that this rectification process is exactly what we need to eliminate vertical disparity in stereoscopic movies, so that all the available results from Computer Vision on computing the epipolar geometry [24, 18, 43, 6, 58] or on the rectification of stereoscopic pairs [23, 39, 20, 1, 74, 80, 14] will be useful to accurately eliminate vertical disparity.

However, the rectification of a single image pair acquired in a laboratory with infinite depth of field for stereoscopic matching, and rectification of a stereoscopic movie that will be presented to spectators is not exactly the same task. The requirements of a rectification method designed for stereoscopic cinema are:

1. it should be able to work without knowing anything from the stereoscopic camera parameters, because this data is not always available, and may be lost in the film production pipeline, whereas images are usually not lost;
2. it should require no calibration pattern or grid, since the wide range of camera configurations and optics would require too many different calibration grid sizes: all the computation should be made from the images themselves (we call it *blind* rectification);
3. the aspect ratio of rectified images should be as close as possible to the aspect ratio of the original images;
4. the rectified image should fill completely the frame: no “unknown” or “black” area can be tolerated;
5. the rectification of the whole movie should be smooth (jitter or fast-varying rectification transforms will create shaky movies) so that the rectification parameters should either be computed for a whole shot, or it should be slowly varying over time;
6. the camera parameters (focal length, focus...) and the rig parameters (interocular, vergence..) may be fixed during a shot, but they could also be slowly varying;

7. the images may have artistic qualities that are difficult to handle for computer vision algorithms: lack of texture, blur, saturation (white or black), sudden change in illumination, noise, etc.

Needless to say, very few rectification methods fulfill these requirements, and the best solution will probably have to take the best out of the cited methods in order to give acceptable results. Cheng et al [14] recently proposed a method to rectify stereo sequences with varying camera motions and zooming effects, while keeping the image aspect ratio, but there is still room for improvements, and there will probably be many other publications on this subject in the near future.

A proper solution would probably contain the following ingredients:

- A method for epipolar geometry computation and rectification that takes into account nonlinear distortion [6, 58, 1]: even if cinema optics are close-to-perfect, images have a very high resolution, and small nonlinear distortions with an amplitude of a few pixels may happen at short focal lengths.
- A multi-scale feature detection and matching method, which will be able to handle blurred or low-texture areas in the images.
- Proper parameterization of the rectification functions, so that only rectifications that conserve the aspect ratio are allowed. Methods for panorama stitching reached this goal by using the camera rotation around its optical center as a parameter – this may be a good direction.
- Temporal filtering, which is not an easy task since the filtered rectifications must also satisfy the above constraints.

5.2 Shifting and Scaling the Images

Image shifting is a process already used by Spottiswoode et al [57], in particular to achieve *continuity in space* during shot transitions. The human brain is actually not very sensitive to stereoscopic shifting: if two fixed rectified images are shown in stereo to a subject, and the images are shifted slowly, the subject will not notice the change in shift, and surprinsingly the scene will not appear to move closer or farther as would be expected. However, shifting the images modifies the eye vergence and not the accomodation, and thus may break the vergence-accomodation constraints and cause visual fatigue: when images are shifted, one must verify that the disparities are still within Percival's zone of comfort (which depends on viewing geometry).

Image shifting is mostly used for "softening the cuts": If the disparity of the main area of interest changes abruptly from one shot to the other, the human eyes may take some time to adjust the vergence, and it will cause visual fatigue or even alter the oculo-motor system, as shown by Emoto et al [17]. As explained by Spottiswoode et al [57], shifting the images is one solution to this problem.

One way to make a smooth transition between shots is the following: let us say the main subject of shot 1 is at disparity d_1, and the main subject of shot 2 is at

disparity d_2 (they do not have to be at the same position in the image). About one second before the transition, the images from shot 1 are slowly shifted from $d_0 = 0$ to $d_0 = \frac{d_2 - d_1}{2}$. When the transition (cut or fade) happens, the disparity of the main subject in shot 1 is $d = \frac{d1+d2}{2}$. During the transition, the first image of shot 2 is presented with a shift of $d_0 = \frac{d1-d2}{2}$, so that the disparity of the main subject is also $d = \frac{d1+d2}{2}$. After the transition, the disparity continues to be slowly shifted, in order to arrive at a null shift ($d_0 = 0$) about one second after the transition.

During this process, care must be taken not only to stay inside Percival's zone of comfort, but also to avoid divergence in the areas in focus (the divergence threshold depends on screen size).

Another simple post-production process consists in scaling the images. Scaling is equivalent to changing the width W of the convergence plane, and allows reframing the scene by panning simultaneously both rescaled images.

There is no quality loss when shifting or scaling the images, except the one due to resampling the original images. In particular, no spatial or temporal artifact may appear in shifted or scaled sequences, whereas they may happen in the processes described hereafter.

5.3 View Interpolation, View Synthesis, and Disparity Remapping

5.3.1 Definitions and Existing Work

Image shifting and scaling cannot solve all the issues caused by having different (i.e. non-homothetic) shooting and viewing geometries. Ideally, we would like to be able to change two parameters of the shooting geometry in order to get homothetic setups: the camera interocular b and the distance to the convergence plane H. This section describes the computer vision and computer graphics techniques which can be used to achieve these effects in stereoscopic movie post-production. The results that these methods have obtained so far are not on par with the picture quality requirements of the stereoscopic cinema, but this research field is progressing constantly, and these requirements will probably be met within the next few years.

The first thing we would like to do is change the camera interocular b, in order to follow the depth consistency rule $\frac{b}{H} = \frac{b'}{H'}$. To achieve this, we use *view interpolation*, a technique that takes as input a set of synchronized images taken from different but close viewpoints, and generates a new view *as if* it were taken by a camera placed at a specified position between the original camera positions. It is different from a well-known post-production effect called *retiming*, which generates intermediate images between two consecutive images in time taken by the same camera, although some people have been successfully using retiming techniques to do view interpolation if the cameras are close enough. As shown in the first line of Fig. 11, 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 change the distance to screen, we have to use *view synthesis*. It is a similar technique, where the cameras may be farther apart (they can even surround the scene), and the synthesized viewpoint can be placed more freely. View synthesis usually requires a shooting setup with at least a dozen of cameras, and may even use hundreds of cameras. It is sometimes used in free-viewpoint video, when the scene is represented as a set of images, not as a textured 3-D mesh. The problem is that we usually only have two cameras, and for view synthesis to work, at least all parts of the scene that would be visible in the synthesized viewpoint must be also visible in the original viewpoints. Unfortunately, we easily notice (second line of Fig. 11) that many parts of the scene that were not visible in the original viewpoints become visible in the interpolated viewpoints (for example the tree in the figure). Since invisible parts of the scene cannot be invented, view synthesis is clearly not the right technique to solve this problem.

What we propose is a mixed technique between view interpolation and view synthesis, that preserves the global visibility of objects in the original viewpoints, but does not produce depth distortion or divergence: *disparity remapping*. In disparity remapping (third line of Fig. 11), 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). In practice, it consists in shifting all the pixels in the image that have a given disparity by the disparity value that would be perceived in the viewing geometry for an object placed at the same depth. That way, divergence may not happen, since objects that were at a given distance from the convergence plane will be projected at the same distance, up to a fixed scale factor, and thus points at infinity are effectively projected at infinity. However, since the object image size is not changed by disparity remapping, there will still be some kind of puppet-theater effect, since far-away objects may appear bigger in the image than they should be. Of course, any disparity remapping function, even one that does not conserve depth, could be used to get special effects, to obtain effects similar to the multi-rigging techniques used in animation (Sec. 4.4).

There has been already a lot of work on view interpolation [81, 15, 46, 64, 10, 76, 82, 25, 32, 72, 75, 68, 50], and disparity remapping as we defined it uses the exact same tools. They have solved the problem with various levels of success, and the results are constantly improving, but none could definitely get rid of the two kinds of artifacts that plague view interpolation:

- spatial artifacts, which are usually "phantom" objects or surfaces that appear due to specular reflections, occluded areas, blurry or non-textured areas, semi-transparent scene components (although many recent methods handle semi-transparency in some way).
- temporal artifacts, which are mostly "blinking" effects that happen if the spatial artifact are not temporally consistent and seem to "pop up" at each frame.

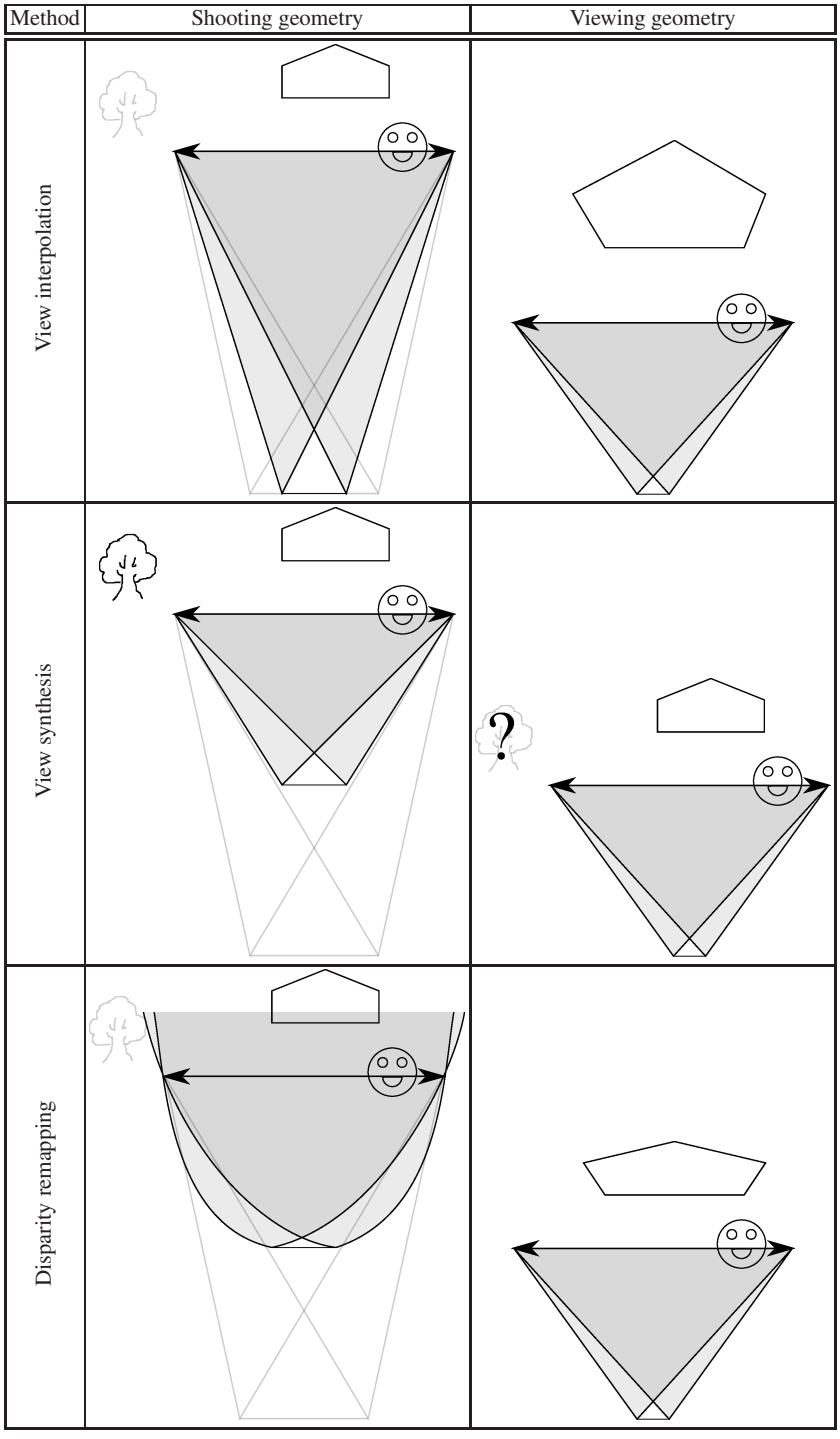


Fig. 11 Changing the shooting geometry in post-production (original shooting geometry is drawn in light gray in the left column). All methods can preserve the roundness of objects in the screen plane, at some cost for out-of-screen objects: view interpolation distorts dramatically their depth and size; view synthesis generates a geometrically correct scene, but it is missing a lot of scene information about objects that cannot be seen in the original images; disparity remapping may preserve the depth information of the whole scene, but the apparent width of objects is not preserved, resulting in puppet-theater effect.

5.3.2 Asymmetric Processing

The geometry changes leading to view interpolation shown in Fig. 11 are symmetric, i.e. the same changes are made to the left and right view. Interpolating both views may alter the quality and create artifacts and both views, resulting in a lower-quality stereoscopic movie. However, this may not be the best way to change the shooting geometry by interpolation.

Stelmach et al [59] showed that the quality of the binocular percept is dominated by the sharpest image. Consequently, the artifacts could be reduced by interpolating only one of the two views, and perhaps smoothing the interpolated view if it contains artifacts. However, a more recent study by Seuntiens et al [51] contradicts the work of Stelmach et al [59], and shows that the perceived-quality of an assymetrically-degraded image pair is about the average of both perceived qualities when one of the two views is degraded by a very strong JPEG compression. Still, they admit that "asymmetric coding is a valuable way to save bandwidth, but one view must be of high quality (preferably the original) and the compression level of the coded view must be within acceptable range", so that for any kind of assymetric processing of a stereo pair (such as view interpolation), one has to determine what is this "acceptable range" to get a perceived quality which is better than with symmetric processing. Kooi and Toet [33] also note that interocular blur suppression, i.e. fusing an image with a blurred image, requires a few more seconds than with the original image pair. Gorley and Holliman [21] devised another metric for stereoscopic pair quality, which is based on comparing areas around SIFT features. This proved to be more consistent with perceived quality measured by human subjects than the classical PSNR (Peak Signal-to-Noise Ratio) measurement used in 2-D image quality measurements. Further experiments should be led on the perception of assymetrically-coded stereoscopic sequences to get a better idea of the allowed amount of degradation in one of the images.

One could argue that the best-quality image in an asymmetrically-processed stereoscopic sequence should be the one corresponding to the dominant eye[5], but this means that interpolation should depend on the viewer, which is not possible in a movie theater or in consumer 3DTV. Seuntiens et al [51] also studied the effect of eye dominance on the perception of assymetrically-compressed images, and their conclusion is that eye dominance has no effect on the quality perception, which means that it may be possible to do assymetric processing while remaining user-independent.

[5] About seventy per cent of people prefer the right eye, and thirty per cent prefer the left, and this is correlated to right- or left-handedness, but many people are cross-lateral (e.g. right-handed with a left dominant eye).

5.4 Changing the Depth of Field

We have seen in sections 2.5.3 and 3.5 that not only the shooting geometry should be adapted to the display, but also the depth of field: the in-focus plane should be at the screen depth (i.e. at zero disparity), and the depth of field should follow the human depth of field (i.e. 0.2D to 0.3D), so that there is no visual fatigue due to horizontal disparity limits or vergence-accomodation conflicts.

Having a limited depth of field may seem limited at first, but Ukai and Howarth [67] showed that 3-D is actually perceived even for blurred objects, and this also results in less visual fatigue when these objects are out of the theoretical focus range. Another big issue with stereoscopic displays is the presence of crosstalk. Lenny Lipton claims [38]: "Working with a digital projector, which contributes no ghosting artifact, proves that the presence of even a faint ghost image may be as important as the accomodation/convergence breakdown.". Actually, blurring the objects that have high disparities should also strongly reduce that ghosting effect, so that changing the depth of field may have the double advantage of diminishing visual fatigue and reducing crosstalk.

In animated 3-D movies, the depth of field can be artificially changed, since images with infinite depth of field can be generated, together with a depth map: [31, 7, 36]. In live-action stereoscopic cinema, the depth of field can be reduced using similar techniques, given a depth map extracted from the live stereoscopic sequence. Increasing the depth of field is theoretically possible, but would enhance dramatically the image noise.

5.5 Dealing With the Proscenium Rule

The proscenium rule states that the stereoscopic sequence should be seen at though it were happening behind a virtual arch formed by the stereoscopic borders of the screen, called the proscenium arch. When an object appears in front of the proscenium, and it touches either the left or the right edges of the screen, then the proscenium rule is violated, because part of the object becomes not visible in both eyes, whereas it should be (see Sec. 4.1 and Fig. 5). When the object touches the top or bottom borders, the stereoscopic scene is still consistent with the proscenium arch, and some even argue that in this case the proscenium appears bended towards the object [42, Chap. 5].

Since the proscenium arch is a virtual window formed by the 3-D reconstruction of the left and right stereoscopic image borders, it can be virtually moved forward or backward. Moving the proscenium forward is the most common solution, especially when using the screen borders as the proscenium would break the proscenium rule, both in live-action stereoscopic movies and in animated 3-D (Sec. 4.3). It can also be a narrative element by itself, as may be learned from animated 3-D (Sec. 4.2). The left and right proscenium edges may even have different depths, without the

audience even noticing it (this effect is called "floating the stereoscopic window" by Mendiburu [42]).

The proscenium arch is usually moved towards the spectator by adding a black border on the left side of the right view and on the right side of the left view. However, on display devices with high rates of crosstalk, moving the proscenium may generate strong artifacts on the screen left and right borders, where one of the views is blackened out. One way to avoid the crosstalk issues is to use a semi-opaque mask instead of a solid black mask, as suggested and used by stereographer Phil Streather. Other alternative solutions solutions may be tried, such as blurring the edges, or a combination of blurring and intensity attenuation.

5.6 Compositing Stereoscopic Scenes and Relighting

Compositing consists in combining several image sources, either from live action or computer-generated, to obtain one shot that is ready to be edited. It can be used to insert virtual objects in real scenes, real objects in virtual scenes, or real objects in real scenes. The different image sources should be acquired using the same camera parameters. For live-action footage, this does usually mean that either match-moving or motion-controlled cameras should be used to get the optical camera parameters (internal and external). Matting masks then have to be extracted for each image source: masks are usually automatically-generated for computer-generated images, and several solutions exist for live-action footage: automatic (blue screen / green screen / chroma key), or semi-automatic (rotoscoping / hand drawing). The image sources are then combined together with matting information to form the final sequence. In non-stereoscopic cinema, if there is no camera motion, the compositing operation is often 2-D only, and a simple transform (usually translation and scaling) is applied to each 2-D image source.

The 3-D nature of the stereoscopic cinema imposes a huge constraint on compositing: not only the image sources must be placed correctly in each view, but the 3-D geometry must be consistent across all image sources. In particular, this means that the shooting geometry (Sec. 3) must be the same, or must be adjusted in post-production to match a reference shooting geometry (Sec. 5.3). If there is a 3-D camera motion, it can be recovered using the same match-moving techniques as in 2-D cinema, which may take into account the rigid link between left and right cameras to recover a more robust camera motion.

The matting stage could also be simplified in stereoscopic cinema: the disparity map (which gives the horizontal disparity value at each pixel in the left and right rectified views) or the depth map (which can be easily computed from the disparity map) can be used as a Z-buffer to handle occlusions automatically between the composited scenes. The disparity map representation should be able to handle transparency, since the pixels at depth discontinuities should have two depths, two colors, and a foreground transparency value [81]. Recent algorithms for stereo-based 3-D reconstruction automatically produce both a disparity map and

matting information, which can be used in the compositing and view interpolation stages [75, 64, 10, 76, 53]. When working on a sequence that has to be modified using view interpolation or disparity remapping, it should be noted that the interpolated disparity map can be obtained as a simple transform of the disparity map computed from the original rectified sequence, and this solution should be preferred to avoid the 3-D artifacts caused by running a stereo algorithm on interpolated images (more complicated transforms happen in the case of view synthesis).

Getting consistent lighting between the composited image sources can be difficult: even if the two scenes to be composed were captured or rendered using the same lighting conditions, the mutual lighting (or radiosity) will be missing from the composited scene, as well as the cast shadows from one scene to the other.

If lighting consistency cannot be achieved during the capture of real scenes (for example if one of the shots comes from archive footage), the problem becomes much more complicated: it is very difficult to change the lighting of a real scene, even when its approximate 3-D geometry can be recovered (by stereoscopic reconstruction in our case), because that 3-D reconstruction will miss the material properties of the 3-D surfaces, i.e. how each surface reacts to incident light, and how light is back-scattered from the surface, and the 3-D texture and normals of the surface are very difficult to estimate from stereo. These material properties are contained in the BRDF (Bi-Directional Reflectance Function), which has a complicated general form with many parameters, but can be simplified from some classes of materials (e.g. isotropic, non-specular...).

The brute force solution to this problem would be to acquire the real scene under all possible lighting conditions, and then select the right one from this huge data set. Hopefully, only a subset of all possible lighting conditions is necessary to be able to recompose any lighting condition: A simple set of lighting conditions (the light basis) is used to capture the scene, and more complicated lighting conditions can be computed by linear combination of the images taken under the light basis. This is the solution adopted by Debevec et al. for their *Light Stage* setup [13, 16].

5.7 Adding Titles or Subtitles

Whereas adding titles or subtitles in 2-D is rather straightforward (the titles should not overlap with important information in the image, such as faces or text), it becomes more problematic in stereoscopic movies.

Probably the most important constraint is that titles should never appear inside or behind objects in the scene: this would result in a conflict between depth (the object is in front of the title) and occlusion (the text occludes the object) which will cause visual fatigue in the long run.

Another constraint is that titles should appear at the same depth as the focus of attention, because re-adjusting the vergence may take as much as one second, and cause visual fatigue [17]. The position of the title within the image is less important, since the oculomotor system has a well-trained “scanning” function for looking at

different places at the same depth. If the display device has too much crosstalk, then a better idea may be to find a place within the screen plane which is not occluded by objects in the scene (because they are highly contrasted, out-of-screen titles may cause double images in this case).

A gross disparity map produced by automatic 3-D reconstruction should be enough to help satisfy these two constraints, the only additional information that has to be provided is *where* is the person who is speaking. Then, the position and depth can be computed from the disparity map and the speaker location: The position can be at the top or the bottom of the image, or following the speaker's face like a speech ballon, and the depth should either be the screen depth (at zero disparity) or the depth of the speaker's face.

6 Conclusion

This chapter has reviewed the current state of understanding in stereoscopic cinema. It has discussed perceptual factors, choices of camera geometry at time of shooting, and a variety of post-production tools to manipulate the the 3-D experience. We included lessons from animated stereoscopy to illustrate how the creative process proceeds when there is complete control over camera geometry and 3-D content, indicating useful goals for stereoscopic live-action work. Looking to the future, we anticipate that 3-D screens will become pervasive in cinemas, in the home as 3-D TV, and in hand-held 3-D displays. A new market in consumer stereoscopic photography has also made an appearance in 2009, as Fuji released the W1 Real 3-D binocular-stereoscopic camera plus auto-stereoscopic photo frame. All of this activity promises exciting developments in stereoscopic viewing and in the underlying technology of building sophisticated 3-D models of the world geared towards stereoscopic content.

References

[1] Abraham S, Förstner W (2005) Fish-eye-stereo calibration and epipolar rectification. ISPRS Journal of Photogrammetry and Remote Sensing 59(5):278–288, DOI 10.1016/j.isprsjprs.2005.03.001 5.1, 5.1

[2] Akeley K, Watt SJ, Girshick AR, Banks MS (2004) A stereo display prototype with multiple focal distances. ACM Trans Graph 23(3):804–813, DOI 10.1145/1015706.1015804 2.5.3

[3] Allison RS (2004) The camera convergence problem revisited. In: Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XI, vol 5291, pp 167–178, DOI 10.1117/12.526278, URL http://www.cse.yorku.ca/percept/papers/Allison-Cameraconvergenceproblemrevisited.pdf 2.5.2

[4] Allison RS (2007) Analysis of the influence of vertical disparities arising in toed-in stereoscopic cameras. Journal of Imaging Science and Technology 51(4):317–327, URL http://www.cse.yorku.ca/percept/papers/jistpaper.pdf 2.5.2

[5] Allison RS, Rogers BJ, Bradshaw MF (2003) Geometric and induced effects in binocular stereopsis and motion parallax. Vision Research 43:1879–1893, DOI 10.1016/S0042-6989(03)00298-0, URL http://www.cse.yorku.ca/percept/papers/Allison-Geometric_and_induced_effects.pdf 2.5.2

[6] Barreto JP, Daniilidis K (2005) Fundamental matrix for cameras with radial distortion. In: Proc. ICCV, DOI 10.1109/ICCV.2005.103, URL http://www.cis.upenn.edu/~kostas/mypub.dir/barreto05iccv.pdf 5.1, 5.1
[7] Bertalmio M, Fort P, Sanchez-Crespo D (2004) Real-time, accurate depth of field using anisotropic diffusion and programmable graphics cards. In: Proc. 2nd Intl. Symp. on 3D Data Processing, Visualization and Transmission (3DPVT), pp 767–773, DOI 10.1109/TDPVT.2004.1335393 5.4
[8] Blake A, Bülthoff H (1990) Does the brain know the physics of specular reflection? Nature 343(6254):165–168, DOI 10.1038/343165a0 3.6
[9] Blake A, Bulthoff H (1991) Shape from specularities: Computation and psychophysics. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 331(1260):237–252, DOI 10.1098/rstb.1991.0012 3.6
[10] Bleyer M, Gelautz M, Rother C, Rhemann C (2009) A stereo approach that handles the matting problem via image warping. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), DOI 10.1109/CVPRW.2009.5206656, URL http://research.microsoft.com/pubs/80301/CVPR09_StereoMatting.pdf 5.3.1, 5.6
[11] Bordwell D (2009) Coraline, cornered. URL http://www.davidbordwell.net/blog/?p=3789, online, accessed 10-Jun-2009, archived at http://www.webcitation.org/5nFFuaq3f 2.3
[12] Campbell FW (1957) The depth of field of the human eye. Journal of Modern Optics 4(4):157–164, DOI 10.1080/713826091 2.5.3
[13] Chabert CF, Einarsson P, Jones A, Lamond B, Ma WC, Sylwan S, Hawkins T, Debevec P (2006) Relighting human locomotion with flowed reflectance fields. In: SIGGRAPH '06: ACM SIGGRAPH 2006 Sketches, ACM, New York, NY, USA, p 76, DOI 10.1145/1179849.1179944, URL http://gl.ict.usc.edu/research/RHL/SIGGRAPHsketch_RHL_0610.pdf 5.6
[14] Cheng CM, Lai SH, Su SH (2009) Self image rectification for uncalibrated stereo video with varying camera motions and zooming effects. In: Proc. IAPR Conference on Machine Vision Applications (MVA), Yokohama, Japan, URL http://www.mva-org.jp/Proceedings/2009CD/papers/02-03.pdf 5.1, 5.1
[15] Criminisi A, Blake A, Rother C, Shotton J, Torr PH (2007) Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming. Int J Comput Vision 71(1):89–110, DOI 10.1007/s11263-006-8525-1 5.3.1
[16] Debevec P, Wenger A, Tchou C, Gardner A, Waese J, Hawkins T (2002) A lighting reproduction approach to live-action compositing. ACM Trans Graph (Proc ACM SIGGRAPH 2002) 21(3):547–556, DOI http://doi.acm.org/10.1145/566654.566614 5.6
[17] Emoto M, Niida T, Okano F (2005) Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. Journal of Display Technology 1(2):328–340, DOI 10.1109/JDT.2005.858938, URL http://www.nhk.or.jp/strl/publica/labnote/pdf/labnote501.pdf 2.5.3, 7, 5.2, 5.7
[18] Fitzgibbon AW (2001) Simultaneous linear estimation of multiple view geometry and lens distortion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol 1, pp 125–132, DOI 10.1109/CVPR.2001.990465, URL http://www.robots.ox.ac.uk/~vgg/publications/papers/fitzgibbon01b.pdf 5.1
[19] Forsyth D, Ponce J (2003) Computer Vision: A Modern Approach. Prentice-Hall 5.1
[20] Fusiello A, Trucco E, Verri A (2000) A compact algorithm for rectification of stereo pairs. Machine Vision and Applications 12:16–22 5.1
[21] Gorley P, Holliman N (2008) Stereoscopic image quality metrics and compression. In: Woods AJ, Holliman NS, Merritt JO (eds) Proc. SPIE Stereoscopic Displays and Applications XIX, SPIE, vol 6803, p 680305, DOI 10.1117/12.763530, URL http://www.dur.ac.uk/n.s.holliman/Presentations/SDA2008_6803-03.PDF 5.3.2
[22] Gosser HM (1977) Selected Attempts at Stereoscopic Moving Pictures and Their Relationship to the Development of Motion Picture Technology, 1852-1903. Ayer Publishing 1.2
[23] Hartley R (1999) Theory and practice of projective rectification. International Journal of Computer Vision 35:115–127 5.1
[24] Hartley R, Zisserman A (2000) Multiple-View Geometry in Computer Vision. Cambridge University Press 5.1
[25] Hasinoff SW, Kang SB, Szeliski R (2006) Boundary matting for view synthesis. Comput Vis Image Underst 103(1):22–32, DOI 10.1016/j.cviu.2006.02.005, URL http://www.cs.toronto.edu/~hasinoff/pubs/hasinoff-matting-2005.pdf 5.3.1
[26] Held RT, Cooper EA, O'Brien JF, Banks MS (2009) Making big things look small: The effect of blur on perceived scale, submitted to ACM Trans. Graph. 3
[27] Held RT, Cooper EA, O'Brien JF, Banks MS (2010) Using blur to affect perceived distance and size. ACM Trans Graph 29(2):1–16, DOI http://doi.acm.org/10.1145/1731047.1731057 2.1
[28] Hoffman DM, Girshick AR, Akeley K, Banks MS (2008) Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. Journal of Vision 8(3):1–30, DOI 10.1167/8.3.33 2.1, 2.5.3, 2.5.3
[29] Hummel R (2008) 3-D cinematography. American Cinematographer pp 52–63 1.3.1
[30] Jones GR, Holliman NS, Lee D (2004) Stereo images with comfortable perceived depth. US Patent 6798406, URL http://www.google.com/patents?vid=USPAT6798406 2.5.1
[31] Kakimoto M, Tatsukawa T, Mukai Y, Nishita T (2007) Interactive simulation of the human eye depth of field and its correction by spectacle lenses. Computer Graphics Forum 26(3):627–636, DOI 10.1111/j.1467-8659.2007.01086.x, URL http://nis-lab.is.s.u-tokyo.ac.jp/~nis/abs_eg.html 5.4

[32] Kilner J, Starck J, Hilton A (2006) A comparative study of free-viewpoint video techniques for sports events. In: Proc. 3rd European Conference on Visual Media Production, London, UK, pp 87–96, URL http://www.ee.surrey.ac.uk/CVSSP/VMRG/Publications/kilner06cvmp.pdf 5.3.1
[33] Kooi FL, Toet A (2004) Visual comfort of binocular and 3D displays. Displays 25(2-3):99–108, DOI 10.1016/j.displa.2004.07.004 2.5, 5.3.2
[34] Kozachik P (2009) 2 worlds in 3 dimensions. American Cinematographer 90(2):26 2.3
[35] Lambooij MTM, IJsselsteijn WA, Heynderickx I (2007) Visual discomfort in stereoscopic displays: a review. In: Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XIV, vol 6490, DOI 10.1117/12.705527 2.5, 2.5.1
[36] Lin HY, Gu KD (2007) Photo-realistic depth-of-field effects synthesis based on real camera parameters. In: Advances in Visual Computing (ISVC 2007), Lecture Notes in Computer Science, vol 4841, Springer-Verlag, pp 298–309, DOI 10.1007/978-3-540-76858-6_30 5.4
[37] Lipton L (1982) Foundations of the Stereoscopic Cinema. Van Nostrand Reinhold 1.2, 2.1, 2.1, 2.2, 3.1
[38] Lipton L (2001) The stereoscopic cinema: From film to digital projection. SMPTE Journal pp 586–593 5.4
[39] Loop C, Zhang Z (1999) Computing rectifying homographies for stereo vision. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol 1, pp –131 Vol. 1, DOI 10.1109/CVPR.1999.786928, URL http://research.microsoft.com/~zhang/Papers/TR99-21.pdf 5.1
[40] Marcos S, Moreno E, Navarro R (1999) The depth-of-field of the human eye from objective and subjective measurements. Vision Research 39(12):2039–2049, DOI 10.1016/S0042-6989(98)00317-4 2.1
[41] Masaoka K, Hanazato A, Emoto M, Yamanoue H, Nojiri Y, Okano F (2006) Spatial distortion prediction system for stereoscopic images. Journal of Electronic Imaging 15(1), DOI 10.1117/1.2181178, URL http://www.nhk.or.jp/strl/publica/labnote/pdf/labnote505.pdf 2.4, 3.1
[42] Mendiburu B (2009) 3D Movie Making: Stereoscopic Digital Cinema from Script to Screen. Focal Press 1.2, 1.3.2, 2.5, 3.2, 5.5
[43] Micusik B, Pajdla T (2003) Estimation of omnidirectional camera model from epipolar geometry. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), vol 1, pp 485–490, DOI 10.1109/CVPR.2003.1211393, URL ftp://cmp.felk.cvut.cz/pub/cmp/articles/micusik/Micusik-CVPR2003.pdf 5.1
[44] Ogle KN (1938) Induced size effect: I. a new phenomenon in binocular space perception associated with the relative sizes of the images of the two eyes. American Medical Association Archives of Ophtalmology 20:604–623, URL http://www.cns.nyu.edu/events/vjclub/classics/ogle-38.pdf 2.5.2
[45] Ogle KN (1964) Researches in binocular vision. Hafner, New York 2.5.2
[46] Park JI, Um GM, Ahn C, Ahn C (2004) Virtual control of optical axis of the 3DTV camera for reducing visual fatigue in stereoscopic 3DTV. ETRI Journal 26(6):597–604 5.3.1
[47] Pastoor S (1992) Human factors of 3DTV: an overview of current research at Heinrich-Hertz-Institut Berlin. In: IEE Colloquium on Stereoscopic Television, London, pp 11/1–11/4, URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=193706 2.5, 2.5.1
[48] Read JCA, Cumming BG (2006) Does depth perception require vertical-disparity detectors? Journal of Vision 6(12):1323–1355, DOI 10.1167/6.12.1, URL http://journalofvision.org/6/12/1 4
[49] Rogers BJ, Bradshaw MF (1993) Vertical disparities, differential perspective and binocular stereopsis. Nature 361:253–255, DOI 10.1038/361253a0, URL http://www.cns.nyu.edu/events/vjclub/classics/rogers-bradshaw-93.pdf 2.5.2
[50] Rogmans S, Lu J, Bekaert P, Lafruit G (2009) Real-time stereo-based view synthesis algorithms: A unified framework and evaluation on commodity GPUs. Signal Processing: Image Communication 24(1-2):49–64, DOI 10.1016/j.image.2008.10.005, URL http://www.sciencedirect.com/science/article/B6V08-4TT34BC-3/2/03a35e44fbc3c0f7ff19fcdbe474f73a, special issue on advances in three-dimensional television and video 5.3.1
[51] Seuntiens P, Meesters L, Ijsselsteijn W (2009) Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation. ACM Trans Appl Percept 3(2):95–109, DOI 10.1145/1141897.1141899 5.3.2
[52] Sexton I, Surman P (1999) Stereoscopic and autostereoscopic display systems. IEEE Signal Processing Magazine 16(3):85–99, DOI 10.1109/79.768575 1.1
[53] Sizintsev M, Wildes RP (2010) Coarse-to-fine stereo vision with accurate 3D boundaries. Image and Vision Computing 28(3):352 – 366, DOI 10.1016/j.imavis.2009.06.008, URL http://www.cse.yorku.ca/techreports/2006/CS-2006-07.pdf 5.6
[54] Smith C, Benton S (1983) reviews of Foundations of the stereoscopic cinema by Lenny Lipton. Optical Engineering 22(2), URL http://www.3dmagic.com/Liptonreviews.htm 1.2, 3.1
[55] Smolic A, Kimata H, Vetro A (2005) Development of MPEG standards for 3D and free viewpoint video. Tech. Rep. TR2005-116, MERL, URL http://www.merl.com/papers/docs/TR2005-116.pdf 1.1
[56] Speranza F, Stelmach LB, Tam WJ, Glabb R (2002) Visual comfort and apparent depth in 3D systems: effects of camera convergence distance. In: Proc. SPIE Three-Dimensional TV, Video and Display, SPIE, vol 4864, pp 146–156, DOI 10.1117/12.454900 2.5.2
[57] Spottiswoode R, Spottiswoode NL, Smith C (1952) Basic principles of the three-dimensional film. SMPTE Journal 59:249–286, URL http://www.archive.org/details/journalofsociety59socirich 1.2, 3.1, 3.2, 3.2, 5.2
[58] Steele RM, Jaynes C (2006) Overconstrained linear estimation of radial distortion and multi-view geometry. In: Proc. ECCV, DOI 10.1007/11744023_20 5.1, 5.1

[59] Stelmach L, Tam W, Meegan D, Vincent A, Corriveau P (2000) Human perception of mismatched stereoscopic 3d inputs. In: International Conference on Image Processing (ICIP), vol 1, pp 5–8, DOI 10.1109/ICIP.2000.900878 5.3.2

[60] Stelmach LB, Tam WJ, Speranza F, Renaud R, Martin T (2003) Improving the visual comfort of stereoscopic images. In: Proc. SPIE Stereoscopic Displays and Virtual Reality Systems X, vol 5006, pp 269–282, DOI 10.1117/12.474093 2.5.2

[61] Stereographics Corporation (1997) The Stereographics Developer's Handbook - Background on Creating Images for CrystalEyes® and SimulEyes®. URL http://www.reald-corporate.com/scientific/downloads/handbook.pdf 2.1, 1

[62] Stevenson SB, Schor CM (1997) Human stereo matching is not restricted to epipolar lines. Vision Research 37(19):2717–2723, DOI 10.1016/S0042-6989(97)00097-7 2.5.2

[63] Sun G, Holliman N (2009) Evaluating methods for controlling depth perception in stereoscopic cinematography. In: Proc. SPIE Stereoscopic Displays and Applications XX, vol 7237, DOI 10.1117/12.807136, URL http://www.dur.ac.uk/n.s.holliman/Presentations/SDA2009-Sun-Holliman.pdf 2.2

[64] Taguchi Y, Wilburn B, Zitnick C (2008) Stereo reconstruction with mixed pixels using adaptive over-segmentation. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp 2720–2727, DOI 10.1109/CVPR.2008.4587691, URL http://research.microsoft.com/users/larryz/StereoMixedPixels_CVPR2008.pdf 5.3.1, 5.6

[65] Todd JT (2004) The visual perception of 3D shape. Trends in Cognitive Sciences 8(3):115 – 121, DOI 10.1016/j.tics.2004.01.006 2.1

[66] Todd JT, Norman JF (2003) The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? Perception & Psychophysics 65(1):31–47, URL http://app.psychonomic-journals.org/content/65/1/31.abstract 2.2

[67] Ukai K, Howarth PA (2007) Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. Displays 29(2):106–116, DOI 10.1016/j.displa.2007.09.004 2.5, 6, 2.5.3, 3.5, 5.4

[68] Wang L, Jin H, Yang R, Gong M (2008) Stereoscopic inpainting: Joint color and depth completion from stereo images. Computer Vision and Pattern Recognition, 2008 CVPR 2008 IEEE Conference on pp 1–8, DOI 10.1109/CVPR.2008.4587704 5.3.1

[69] Wann JP, Mon-Williams M (1997) Health issues with virtual reality displays: What we do know and what we don't. ACM SIGGRAPH Computer Graphics 31(2):53–57 2.5, 2.5.3

[70] Wann JP, Rushton S, Mon-Williams M (1995) Natural problems for stereoscopic depth perception in virtual environments. Vision Research 35(19):2731–2736, DOI 10.1016/0042-6989(95)00018-U 2.5

[71] Watt SJ, Akeley K, Ernst MO, Banks MS (2005) Focus cues affect perceived depth. J Vis 5(10):834–862, DOI 10.1167/5.10.7, URL http://journalofvision.org/5/10/7/ 2.1

[72] Woodford O, Reid ID, Torr PHS, Fitzgibbon AW (2007) On new view synthesis using multiview stereo. In: Proceedings of the 18th British Machine Vision Conference, Warwick, vol 2, pp 1120–1129, URL http://www.robots.ox.ac.uk/~ojw/stereo4nvs/Woodford07a.pdf 5.3.1

[73] Woods A, Docherty T, Koch R (1993) Image distortions in stereoscopic video systems. In: Proc. SPIE Stereoscopic Displays and Applications IV, San Jose, California, vol 1915, pp 36–48, DOI 10.1117/12.157041, URL http://www.cmst.curtin.edu.au/publicat/1993-01.pdf 2.5.2, 3.1

[74] Wu HHP, Chen CC (2007) Projective rectification with minimal geometric distortion (13). In: Stolkin R (ed) Scene Reconstruction Pose Estimation and Tracking, I-Tech Education and Publishing, pp 221–242, URL http://intechweb.org/book.php?id=10 5.1

[75] Xiong W, Jia J (2007) Stereo matching on objects with fractional boundary. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI 10.1109/CVPR.2007.383194, URL http://www.cse.cuhk.edu.hk/~leojia/all_final_papers/alpha_stereo_cvpr07.pdf 5.3.1, 5.6

[76] Xiong W, Chung H, Jia J (2009) Fractional stereo matching using expectation-maximization. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(3):428–443, DOI 10.1109/TPAMI.2008.98, URL http://www.cse.cuhk.edu.hk/~leojia/all_final_papers/pami_stereo08.pdf 5.3.1, 5.6

[77] Yamanoue H, Okui M, Okano F (2006) Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images. IEEE Trans on Circuits and Systems for Video Technology 16(6):744–752, DOI 10.1109/TCSVT.2006.875213 2.4, 3.2, 3.2

[78] Yano S, Emoto M, Mitsuhashi T (2004) Two factors in visual fatigue caused by stereoscopic HDTV images. Displays pp 141–150, DOI 10.1016/j.displa.2004.09.002 2.5.1, 2.5.3, 2.5.3

[79] Yeh YY, Silverstein LD (1990) Limits of fusion and depth judgment in stereoscopic color displays. Hum Factors 32(1):45–60 2.5.1

[80] Zhou J, Li B (2006) Rectification with intersecting optical axes for stereoscopic visualization. Proc ICPR 2:17–20, DOI 10.1109/ICPR.2006.986 5.1

[81] Zitnick CL, Kang SB, Uyttendaele M, Winder S, Szeliski R (2004) High-quality video view interpolation using a layered representation. In: Proc. ACM SIGGRAPH, ACM, New York, NY, USA, vol 23, pp 600–608, DOI 10.1145/1015706.1015766, URL http://research.microsoft.com/~larryz/ZitnickSig04.pdf 5.3.1, 5.6

[82] Zitnick CL, Szeliski R, Kang SB, Uyttendaele MT, Winder S (2006) System and process for generating a two-layer, 3D representation of a scene. US Patent 7015926, URL http://www.google.com/patents?vid=USPAT7015926 5.3.1
[83] Zone R (2005) 3-D fimmakers : Conversations with Creators of Stereoscopic Motion Pictures. No. 119 in The Scarecrow Fimmakers Series, Scarecrow Press 1.2, 2.3, 3.2
[84] Zone R (2007) Stereoscopic Cinema and the Origins of 3-D Film, 1838-1952. University Press of Kentucky 1.2