Signal Processing for Stereoscopic and Multi-View 3D Displays

Atanas Boev, Robert Bregovic, and Atanas Gotchev

Abstract Displays which aim at visualizing 3D scenes with realistic depth are known as "3D displays". Due to technical limitations and design decisions, such displays might create visible distortions, which are interpreted by the human visual system as artifacts. This book chapter overviews a number of signal processing techniques for decreasing the visibility of artifacts on 3D displays. It begins by identifying the properties of a scene which the brain utilizes for perceiving depth. Further, operation principles of the most popular types of 3D displays are explained. A signal processing channel is proposed as a general model reflecting these principles. The model is applied in analyzing how visual quality is influenced by display distortions. The analysis allows identifying a set of optical properties which are directly related with the perceived quality. A methodology for measuring these properties and creating a *quality profile* of a 3D display is discussed. A comparative study introducing the measurement results on the visual quality and position of the sweet spots of a number of 3D displays of different types is presented. Based on knowledge of 3D artifact visibility and understanding of distortions introduced by 3D displays, a number of signal processing techniques for artifact mitigation are overviewed. These include a methodology for *passband optimization* which addresses typical 3D display artifacts (e.g. Moiré, fixed-pattern-noise and ghosting), a framework for design of tunable anti-aliasing filters and a set of real-time algorithms for view-point based optimization.

A. Boev (✉) • R. Bregovic • A. Gotchev
Department of Signal Processing, Tampere University of Technology, Tampere, Finland
e-mail: atanas.boev@tut.fi; robert.bregovic@tut.fi; atanas.gotchev@tut.fi

S.S. Bhattacharyya et al. (eds.), *Handbook of Signal Processing Systems*,
DOI 10.1007/978-1-4614-6859-2_1,

1 Introduction

A real-world three-dimensional scene is a rich and complex visual phenomenon processed and perceived only partially by the human vision system (HVS). For a high-quality 3D scene representation it is sufficient to recreate only the perceivable visual information and omit the "visually-unnecessary" features. In this way, less data needs to be processed, while keeping the reproduced scene indistinguishable from the real one. However, due to the technical limitations of today's display devices, some visually important features might be lost, which results in perceptual differences between the visual replica and the real scene and is interpreted by the HVS as the presence of artifacts. By using appropriate signal processing techniques it is possible to make these artifacts less noticeable and achieve more pleasant scene visualization. Correspondingly, this chapter addresses two groups of problems. First, it investigates upon what are the necessary components of a scene that should be preserved and reproduced, and second, how to pre-process the scene representation data in order to minimize the perceived distortion and decrease the visibility of artifacts on a 3D display.

The chapter consists of five sections. Section 2.1 discusses which 3D features are visually important and how these features can be included in a 3D scene representation. In Sect. 2.2 a 3D display classification is presented. The classification is based on the method that each display uses to recreate the stereoscopic image. In Sect. 3.1, the knowledge of display specifics is combined with HVS properties in an attempt to explain the appearance and visibility of artifacts on 3D displays. Section 3.2 discusses which optical properties of a 3D display are important from the visual quality point of view and presents a methodology to measure these properties. These measurements allow one to derive the so-called quality profile of a given 3D display. In Sect. 4 the understanding of artifact visibility and knowledge on optical quality is used for a set of image processing algorithms which aim at a visual optimization of a 3D scene. Section 5 provides some conclusions.

2 Principles of 3D Visualization

2.1 3D Scene Perception and Representation

An ideal 3D display would attempt creating a light field being a *perfect visual replica* of a 3D scene. Such a replica, however, would also include components which are not visible to human eyes. These components can be considered redundant and can be omitted from the scene representation. The result is a *visually-indistinguishable replica* of the scene. Furthermore, the typical display use case does not require the scene to react to external light sources or to allow the observer to walk through object in the scene. Thus, some visual information (e.g. light distribution within scene objects) is unnecessary. Removing this information produces a

redundancy-free replica of the scene. In a typical use case a redundancy-free replica is also a visually indistinguishable representation of the scene under the use case constraints. Failure in creating redundancy-free and visually-indistinguishable replica leads to visible distortions. In order to avoid this one needs to know which light properties are important and which scene features are relevant for perceiving the scene in 3D.

2.1.1 Visual Perception of Depth

Vision in general can be separated into two parts—visual perception and visual cognition. In studies of human vision, visual perception and properties of *early vision* are subjects of anatomy and neurophysiology [1, p. 2, 2], and visual cognition, as a higher level brain function, is a subject of psychology [1, p. 387, 3].

Visual perception involves a number of optical and neural transformations. The eye can change its refractive power in order to focus on objects at various distances. The process is known as *accommodation* and the refractive power is measured in diopters. The light entering the eye is focused onto the *retina* which contains photosensitive receptors tuned to various spectral components (frequencies). The density of the photoreceptors has its maximum close to the optical center of the eye. The area with the highest photoreceptor density is known as the *fovea.* There are four types of photoreceptor cells—rods, L-cones, M-cones and S-cones—which allow detection of light with wavelengths between 370 and 730 nm. The cones can be thought of (to a crude approximation) as sensitive to red, green and blue color components of the light. The rods are responsible for the low-light vision and are generally ignored in HVS modeling. Rather than perceiving continuous spectrum, the HVS encodes the color information as a combination of three color components; the process is known as *color perception*. The combination of the iris controlling the amount of light entering the eye, and the sensitivity adaptation of the retina allow the eye to work over a wide range of intensities (between 10^{-6} and 10^{8} cd/m^2). The eye is sensitive to luminance difference (i.e. contrast) rather than absolute luminance values. This visual property is known as *light adaptation*. However, the HVS has different *contrast sensitivity* for patterns with different density and orientation [1].

The ability to perceive visual information through two distinctive eyes is known as *binocular vision*. The eyes of a human are separated horizontally and have distance between pupils (also known as *interpupilar distance*, IPD) of approximately 65 mm on average [2]. Such positioning allows each eye to perceive the world from a different perspective, as shown in Fig. 1. The luminance, color and contrast perception occur in each eye separately and the visual information is fed through the optical nerve to the so-called *lateral geniculate nucleus* (LGN) [1]. The LGN decorrelates the binocular information and produces a single, fused representation of the scene. The fused image appears as if observed from a point between the eyes and is called *cyclopean image*. The point, which is projected in the fovea of each eye, is known as the *point of convergence*. The observer can control the visual fixation point though the *extraocular muscle system*. If an object is around the point of

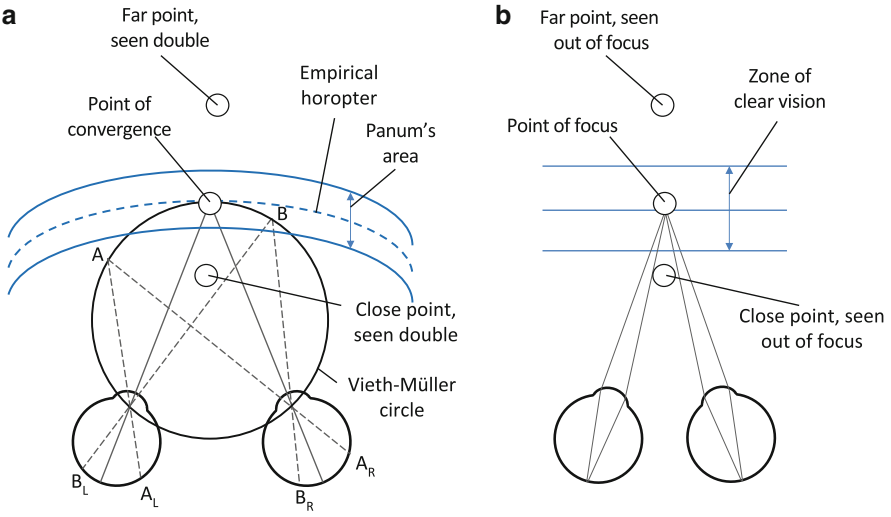


Fig. 1 Binocular geometry: (**a**) horopter for a given point of convergence and (**b**) zone of clear vision for a given point of focus

convergence the HVS can fuse its projections into each eye into a cyclopean image. Since eyes perceive the scene from different perspectives the projections of an object around the point of convergence are not identical. The existence of two different retinal images is called *binocular disparity* [2]. The difference between retinal images allows the brain to deduct information about the relative depth between different points of interest. The ability of the brain to deduct depth information from retinal disparity is known as *stereovision*.

All points that are projected onto identical places in each retina (relative to the fovea) can be fused by the HVS. For a given point of convergence there are points which are projected with identical offset relative to each fovea, as shown for points "A" and "B" in Fig. 1a. The set of all points which are projected onto matching retinal positions is called the *horopter*. The theoretical horopter coincides with the circle which passes through the point of convergence and the center of each eye's lens, as shown in Fig. 1a. That circle is also known as *Vieth-Müller circle*. However, the horopter derived through subjective experiments (also called the *empirical horopter*) does not fully coincide with the theoretical one. Around the horopter there is a region of points at which projections can be fused by the HVS. That region is known as *Panum's area* and outside of this, binocular depth is still perceived but objects are seen as doubled. The experience of seeing double objects is known as *diplopia* [1].

When eyes focus on a point the refractive power of each eye changes in order that the projections of that point appear in focus in each retina, as seen in Fig. 1b. Close to the point of focus there is a larger area where objects are perceived in focus. The area is known as *zone of clear vision* and its size depends on the distance to the point of focus and the size of the iris. In order to speed-up the

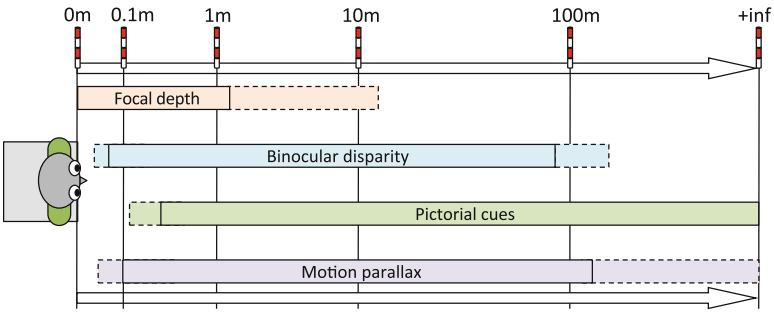


Fig. 2 Depth perception as a set of separate visual "layers"

accommodation process the convergence and focus of the eyes are simultaneously driven by the so-called *accommodation-convergence reflex*. The distance to the point of convergence influences the focal distance, and vice versa. In a natural 3D scene such coupling increases the speed of accommodation and helps the convergence process by blurring the objects in front and behind the convergence point.

Vision in 3D consists of different subsystems which provide separate information about the scene depth. The visual features used by the HVS for perceiving the depth are also known as *depth cues*. There are separate groups of depth cues with varying importance from observer to observer [2, 4]. The presence and strength of one type of depth cue might suppress of enhance the visibility of another. The importance of different depth cues also varies with the distance, as shown in Fig. 2. There are the following groups of depth cues:

- *Focal depth*—The HVS can use the refractive power of the eye as a depth cue. At short distances accommodation is the primary depth cue since closely positioned objects are hardly visible with two eyes. With increasing observation distance the importance of this depth cue quickly drops.
- *Binocular depth*—Retinal disparity is used as a depth cue providing relative distance. Binocular depth cues are the ones most often associated with "3D cinema". Approximately 5% of all people are "stereoscopically latent" and have difficulties assessing binocular depth cues [1, 3]. Such people rely on depth information coming from other cues.
- *Pictorial cues*—for longer distances, binocular depth cues become less important and the HVS relies on pictorial cues for depth assessment such as shadows, perspective lines and texture scaling. Pictorial depth cues can be perceived by a single eye.
- *Head parallax* (also known as *motion parallax*)—this is the process in which the changing parallax of a moving object is used for estimating its depth and 3D shape. Observers naturally expect to be able to see the scene from different perspectives by changing their head position. The same mechanism is used by insects and is commonly known as "insect navigation" [5].

More detailed information about the binocular depth perception can be found in [2, 3, 6].

2.1.2 3D Scene Sensing and Representation

A 3D scene sensing technique attempts to solve the ill-posed problem of reconstructing a 3D scene from a limited number of remote observations. There is a wide range of such techniques, overviewed by Stoykova et al. [7]. One group of methods aims at reconstructing a 3D scene which is captured by a single camera; these methods work by analyzing monocular depth cues. This category includes; *shape-from-shading* [8], *shape-from-texture* [9], *shape-from-defocus* [10] and *shape-from-motion* [11]. Another single-camera 3D sensing approach involves fitting a 3D model over known 3D shapes such as the face [12] or body [13]. This is equivalent to process in which the HVS assumes size and 3D shape of known objects. The second group of techniques attempts to reconstruct a scene captured by two or more cameras. The main problems of these approaches are finding corresponding features in each observation and reconstruction of occluded pixels [14]. The third group of methods uses active camera sensing and captures 3D data by projecting *structured patterns* or *coded light*. Another active 3D sensing approach is *time-of-flight* imaging where the camera emits a light signal and measures the time it takes for the signal to reach the scene and bounce back to the camera [15]. Finally, there are holographic 3D scene capture methods which record the interference pattern created by superimposing a reference beam with a beam scattered by the scene. If the interference pattern is captured by a charge coupled device (CCD) camera instead of holographic material the technique is known as *digital holography* [16].

Usually, 3D scene representation format is a trade-off between two things; first, to have of an accurate description of the "important" visual features and second, to have compact description which is suitable for storing and transmission. Most formats for representing visual data descend from the human understanding of a natural scene in terms of geometry and texture. However, scene description formats are also greatly influenced by peculiarities of the content creation process. While the concrete details in encoding, compression or file structure might differ, there are three major groups of abstract 3D scene representations [17].

The first is so-called *spatio-perspective volume* where a multiple viewpoints of the same scene are recorded [18]. Such volume is created by capturing or rendering images from different camera perspectives. The camera can move in a 2D plane to capture full scene parallax or along a line to capture horizontal parallax only. Due to the similarity between the images seen from neighboring locations (which is called *perspective coherence* in [18]), the spatio-temporal volume is a description which contains great amount of redundancy. Observations of objects captured by a linearly moving camera appear with linear shifts and impose the so-called *epipolar constraint* [19]. As a consequence, a slice of the spatio-temporal volume parallel to the perspective dimension contains many straight lines, as shown in Fig. 3a. The lines are known as *epipolar lines* [19], and the slice is known as *epipolar plane image* (EPI) [18].

When sliced across a perspective dimension, the volume contains a number of scene observations from different perspective (known as *views*). Such scene representation contains a limited number of these observations and is denoted as a

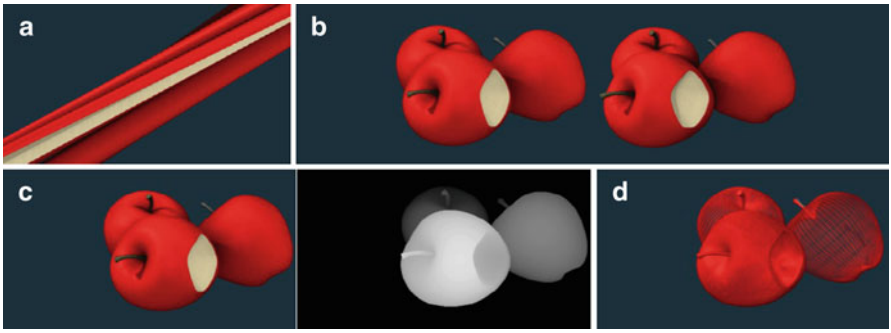


Fig. 3 Representations of a 3D scene: (**a**) epipolar image, (**b**) side-by-side stereoscopic pair, (**c**) 2D + Z image pair, and (**d**) mesh

multiview image [17]. A relatively simple way to store multiview image is to combine all observations in a single bitmap. For a stereoscopic image, both views can be stored in a *side-by-side* fashion as shown in Fig. 3b. A more sophisticated approach is to encode the differences between the observations similarly to the way temporal similarities are encoded in a video file as done in MPEG-4 MVC [20]. Multiview images are one of the most often used formats for natural scene description.

The second group of scene representations is *video-plus-depth* where each pixel is augmented with information of its distance from the camera. A straightforward way to represent video-plus-depth is to encode the depth map as a grey scale picture and place the 2D image and its depth map side-by-side. The intensity of each depth map pixel represents the depth of the corresponding pixel from the 2D image. Such format is sometimes referred to as $2D + Z$ and an example of this representation of a scene is shown in Fig. 3c. Video-plus-depth format can be used to render virtual views based on the geometrical information about the scene encoded in the depth map. Thus, it is suitable for multiview displays and can be used regardless of the number of views a particular screen provides [17, 21]. Furthermore, video-plus-depth can be efficiently compressed. Recently, MPEG specified a container format for video-plus-depth data known as MPEG-4 Part-3 [20]. On the downside, rendering scene observations using 2D + Z description requires disocclusion filling, which can introduce artifacts. This is being addressed by using layered depth images (LDI) [17] or by multi-video-plus-depth encoding [22]. A dense depth map is not captured directly but can be derived from multiview images (using depth estimation algorithms) or from point cloud data captured by range sensors. In the case of a synthetic 3D scene, obtaining a dense depth map is a straightforward process as solving the occlusions during rendering requires calculation of the distance between camera and each pixel of the image [23].

The third group of representations store scene geometry in a vectorized form. One example is a dynamic 3D mesh [20]. Such representation is suitable for synthetic content since synthetic 3D scenes are described in terms of shapes and

textures. An example of mesh representation is shown in Fig. 3d. More details on 3D scene representation formats can be found in [17, 20].

2.2 3D Displays

Three-dimensional displays are ones which aim to show a visually indistinguishable copy of a real 3D scene. The ideal 3D display would recreate all depth cues of a scene, regardless of their importance or applicability in a particular use scenario. In practice, due to design constraints, only a subset of the depth cues is recreated.

Most often, a display earns its "3D" label by being able to provide separate image for each eye of the observer. In a good stereoscopic pair, objects appear on different horizontal coordinates in each image. The horizontal offset between the observations is known as *display disparity*. When a stereoscopic image is observed, display disparity induces retinal disparity, which in turn creates the stereoscopic illusion of depth. The illusory distance to the object created by the stereoscopic effect is called *apparent depth*. Positive disparity creates apparent depth behind the screen plane and negative disparity creates apparent depth in front of the screen.

Most contemporary 3D displays do not recreate head parallax. Some models can present limited head parallax by casting different images towards a set of observation angles, usually limited to a horizontal head parallax only. Note, that by using head-tracking it is possible to present a scene from different perspectives on a monoscopic display, thus generating head parallax without binocular depth cues [24]. Focal depth cues are very rarely recreated by 3D displays. One exception is the stereo display prototype with multiple focal distances described in [25]. Finally, pictorial depth cues can be recreated by most 2D and 3D displays (volumetric LED cube displays [26] being an exception). More information about various types of 3D displays can be found in [27–30].

2.2.1 Classification

A general taxonomy of 3D displays divides them into three basic types; holographic, volumetric and multiple-image screens [27, 28, 31]. Holographic displays use holographic methods to reconstruct the light field of a scene, volumetric displays attempt to approximate a 3D scene by light elements (voxels) positioned in 3D space and multiple image screens cast a number of different images, each one seen from a different angle. There are two types of multiple-image screens. The first type works by tracking the observer's eyes and utilizes steerable optics to beam different images towards each eye. The second type uses fixed optics and beams a number of different images (called "*views*") in different directions; the directions are selected in such way that the eyes of an observer standing in front of the screen perceive different images. In [27] these two types are said to create *eye-gaze-related image* and *fixed-plane image* correspondingly. The taxonomy in [28] is different;

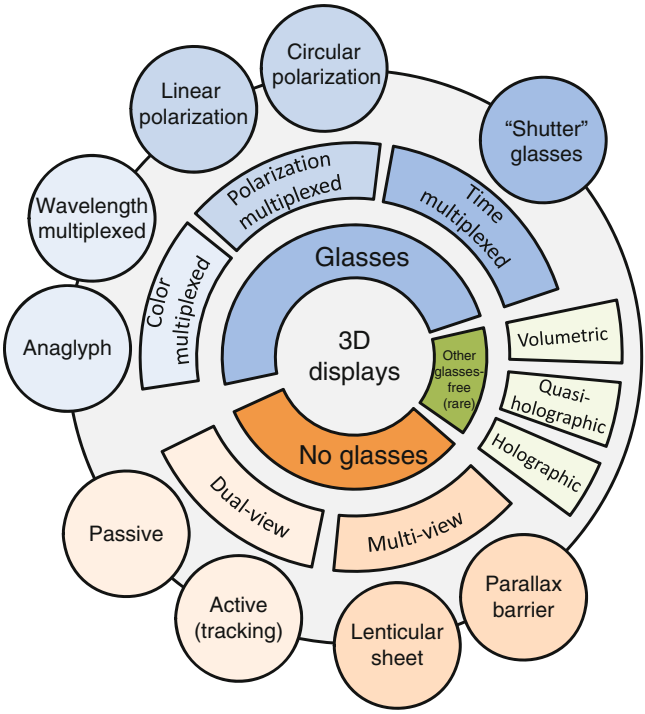


Fig. 4 Classification of 3D displays

displays with steerable optics are named "*head position tracking displays*", while the ones with fixed optics are designated simply as "*multiview displays"*. This chapter follows the terminology in [28] and uses *multiview display* to designate autostereoscopic displays which generate multiple images by means of fixed optics.

The classification used in this book chapter is shown in Fig. 4. It classifies 3D displays from the users' point of view. For the observer, the main difference is whether the display requires glasses or not. Thus the taxonomy in this book chapter has "glasses-based" or" glasses-free" as major display types. The predominant share of 3D displays in the market is binocular stereoscopic TV sets which use thin film transistor liquid crystal displays (TFT–LCD) for image formation and require the observers to wear glasses. Color multiplexed anaglyph glasses are rare, though some 3D cinemas still use wavelength multiplexed glasses [32]. The 3D TV sets are sold either with *polarized glasses* (marketed as "passive") or *temporally-multiplexed* ones (marketed as "active"). The displays without glasses are separated into two groups; binocular autostereoscopic ones mostly used in mobile devices, and multiview displays used for outdoor advertising or (rarely) in computer setups. As an exception, Toshiba announced a 3D TV model which uses a combination of a multiview display and observer tracking [33]. All other types of 3D displays, for example volumetric or holographic ones, are rare and mostly in prototype form.

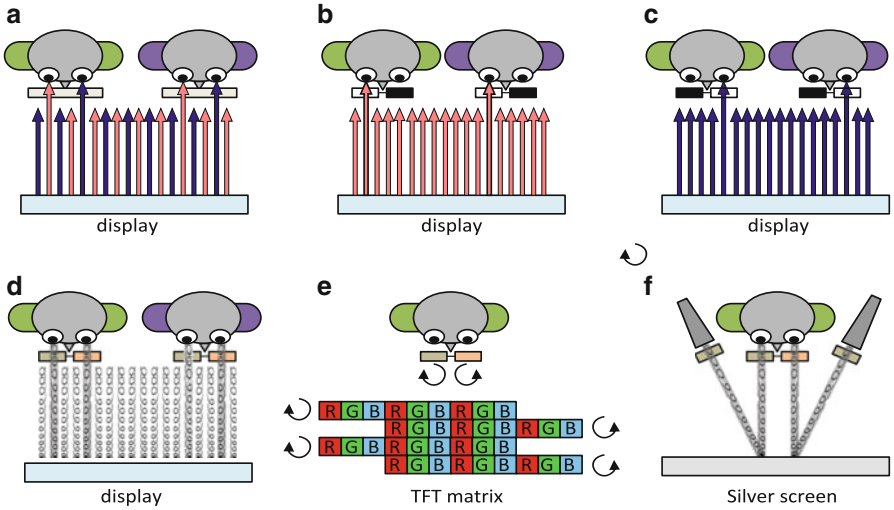


Fig. 5 Glasses-based 3D displays: (**a**) general principle of operation, (**b**, **c**) operation principle of temporary-interleaved glasses, (**b**) left view visible, (**c**) right view visible, (**d**) operation principle of polarization glasses, (**e**) spatially interleaved display using polarized light and (**f**) dual-projection system using polarized light

2.2.2 Glasses-Enabled Stereoscopic Displays

Glasses-enabled 3D displays use one display surface to beam two views (one for each eye). Glasses worn by each observer separate the light beams so each eye receives only the intended view, as shown in Fig. 5a. Temporary-interleaved 3D displays beam both views, alternating them over time. The observer wears active glasses which work synchronously with the display and block the light to one or the other eye at the proper moment. When the display is beaming the left image the light towards the right eye is blocked (Fig. 5b), and when the right image is beamed the light to the right eye is blocked (Fig. 5c). At any moment only one of the observer's eyes perceives the image, but due to the high speed of the process (120–240 frames per second) the user is unaware of the temporal interleaving.

Another approach is to beam both images using differently polarized light and use polarization filters in front of each eye. In this case each eye receives differently polarized light but since the HVS is not sensitive to light polarization the observer is unaware of the separation. Most often, circular polarization is used (clockwise for one eye and counter-clockwise for the other) which allows the beam separation to work for a wide range of head orientations (e.g. head tilt). Passive polarizing glasses are used with both light-emitting TV displays (Fig. 5d) and light-reflecting projector-based displays (Fig. 5f). The light-emitting stereoscopic displays with passive glasses (hereafter denoted by *SDPG*) use spatial interleaving. In such displays the available TFT elements are divided into two groups with different polarization, as shown in Fig. 5e. The groups are usually row-interleaved; the rows

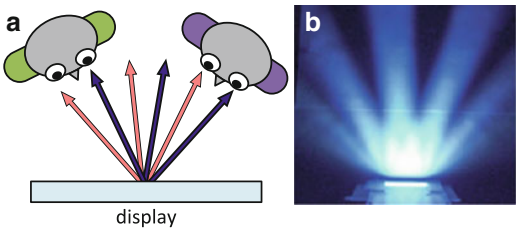


Fig. 6 Dual-view autostereoscopic displays: (**a**) general principle and (**b**) visibility zones for one of the views

with odd numbers are visible by one eye and the rows with even numbers by the other. Each eye sees the other half of the rows dark; for example the left eye may see the image in the odd rows and black stripes in the place of the even rows. Projector-based setups use two projectors equipped with polarizing filters (Fig. 5f) and require a special reflecting surface in order to preserve the polarization of the reflected light. Since two projectors are used, each eye receives image with the same resolution.

2.2.3 Dual-View Autostereoscopic Displays

Dual-view autostereoscopic displays create two views, each one visible from different perspective. Each view is visible from multiple observation angles, as shown in Fig. 6a. This allows a number of observers to use such display provided that each observer is correctly positioned. A practical example of positions where one of the views is visible is shown in Fig. 6b. It shows a photograph of a dual-view autostereoscopic display beaming two images; one "white" image where all pixels are at full brightness, and another "black" where all pixels are off. On the photograph one can see the positions where the "white" view is visible.

There are a number of designs which allow one display to beam two different images. The most common approach is to put an additional layer in front of the TFT–LCD [27, 28, 34]. TFT displays recreate the full color range by emitting light though red, green and blue colored components (*sub-pixels*), usually arranged in repetitive vertical stripes as shown in Fig. 7. The layer alters the visibility of each pixel and makes only half of the sub-pixels visible from a given direction. The layer is called "*optical layer*" [35], "*lens plate*" [27] or "*optical filter*" [36]. The design, where only part of the sub-pixels is visible from a given direction, is also known as *spatially-multiplexed* autostereoscopic display [27].

There are two common types of optical filters, namely; *lenticular sheet* and *parallax barrier*. Lenticular sheets are composed of small lenses which refract the light to different directions as shown in Fig. 7a [35]. A parallax barrier is essentially a mask with openings and closings that block the light in certain directions as shown in Fig. 7b [34]. In both cases the intensity of the light rays passing through the filter changes as a function of the angle as if the light is directionally projected. Also, as only half of the available sub-pixels belong to one of the views, the resolution of each view is lower than the full 2D resolution of the display.

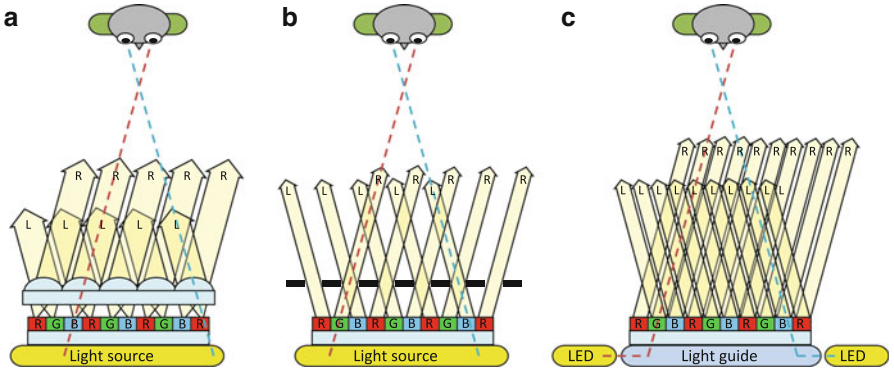


Fig. 7 Optical filters for autostereoscopic displays: (**a**) lenticular sheet, (**b**) parallax barrier and (**c**) temporally-interleaved patterned retarder

One way to have each view with the full resolution of the display is to use temporal interleaving. One example is the 3D display with the patterned retardation film produced by 3M [37]. It distributes the light into two perspective views in a sequential manner, as shown in Fig. 7c. The display uses a standard TFT panel and two separate backlighting sources. The two backlights are turned on and off in counter phase so that each backlight illuminates one view. The switching is synchronized with the LCD which displays different-perspective images at each backlit switch-on time. The role of the 3-D film is to direct the light coming from the activated backlight to the corresponding eye. More information on autostereoscopic displays can be found in [30, 38, 39].

2.2.4 Multiview Displays

Most multiview 3D displays work in a similar fashion to the spatially-multiplexed dual-view ones. However, instead of having their sub-pixels separated into two views, multiview displays have more views, typically 8 to 24. The current generation of multiview displays uses the same basic principles for light distribution; lenticular sheets [35] or slanted parallax barriers [36]. The lenticular sheet works by refracting the light as shown in Fig. 8a, and the parallax barrier works by blocking the light in certain directions as shown in Fig. 8b. In both cases the intensity of the light rays passing through the filter changes as a function of the angle [35]. Since sub-pixels appear displaced in respect to the optical filter, their light is redirected towards different positions. As a result, differently colored components of one pixel belong to different views. Respectively, the image formed by one view will be a combination of color components (sub-pixels) of various pixels across the TFT screen. When red, green and blue sub-pixels are visible from the same direction

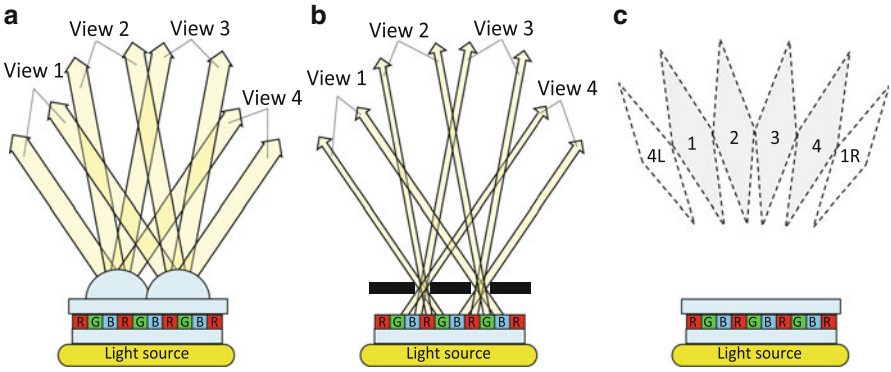


Fig. 8 Multiview displays: (**a**) lenticular sheet, (**b**) parallax barrier and (**c**) visibility zones of the views

and appear close to each other, the triplet is perceived as one full-color pixel. Such pixel is a building block of the view seen from that direction.

As a result of applying the optical filter, for every sub-pixel there is a certain angle from which it is perceived with maximal brightness; that angle we call the *optimal observation angle* for the sub-pixel. The vector, which starts from the sub-pixel and follows the optimal observation angle, is the *optimal observation vector* for the sub-pixel. The optimal observation vectors for all sub-pixels of the same view are designed to intersect in a tight spot in front of the multiview display. From this spot the view will be perceived with its maximal brightness. That spot is referred to as being the *optimal observation spot* of the view. Outside of the optimal observation spot there is a range of observation angles from which a given view is still visible, but with diminished brightness. That range is called the *visibility zone* of a view. For most multiview displays visibility zones of the views are ordered in the horizontal direction. A notable exception is the SynthaGram display produced by StereoGraphics [40] which has nine views with visibility zones ordered in a 3-by-3 grid. As the amount of the pixels provided by the underlying TFT is limited, there is a trade-off between the number of views created by a 3D display and the resolution of each view. As stereoscopic depth cues are perceived mostly in horizontal direction, most multiview display designs do not allocate pixels for extra views in the vertical direction [28, 36, 38, 41].

When horizontally ordered, the visibility zones appear in a fan-shaped configuration as depicted in Fig. 8c. The repetitive structure of the optical filter creates several observation zones for any view; these follow the fan-shaped configuration as well. After the visibility zone of the last view, the first view becomes visible again. This creates one central set of visibility zones directly in front of the screen and a number of identical sets to the side as shown in Fig. 8c. The zones marked as “1” and “1R” are observation zones of the same view.

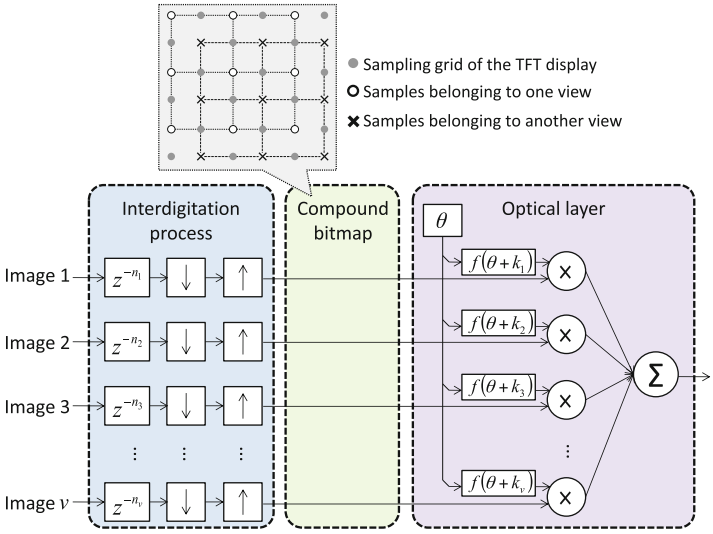


Fig. 9 Model of a multiview display as an image processing channel

2.2.5 Autostereoscopic Displays Modeled as a Signal Processing Channel

In order to relate the optical properties of a 3D display to the visual quality one can consider the display as a signal processing channel. The model has two parts, as shown in Fig. 9. The first part of the model is the process where the sub-pixels of the views are rearranged into one compound bitmap. Such a process is also known as *interdigitation*. The input comes from v images and each image is considered to have the full ("2D") resolution of the display. From each input image, only sub-pixels which belong to one of the views are used. This is modeled by a 2D down-sampling operation. Since the views are spatially-multiplexed, each image gets sampled with different horizontal and vertical offset. On the display the sub-sampled image is represented in its original size. The visible sub-pixels appear either surrounded by black stripes by the parallax barrier, or enlarged by the lenticular sheet. This effect is modeled as an up-sampling stage where the introduced samples are either set to zero, or are repetition of the same sample value.

The optical layer of a multiview display acts as a directionally-selective filter and applies angular luminance function to each sub-pixel of the display. The angle at which the angular luminance has its peak value determines the optimal observation direction of the sub-pixel; this angle is different for each sub-sampled image. The compound bitmap map can be represented as a set of non-overlapping lattices, where each lattice contains sub-pixels from a single view only [40]. On an image with the full resolution of the LCD, each of these lattices acts as a rectangular subsampling pattern with a different offset. The offset is modeled by a signal delay (represented by z-domain blocks in Fig. 9). An example is shown in the top of Fig. 9 where the

intersecting dotted lines mark the position of LCD sub-pixels; one lattice is marked with circles and another is marked with crosses.

The second part of the model represents the effect of the optical layer. The impact of the layer on the brightness of the underlying sub-pixels is modeled as *visibility* which is the ratio between the relative brightness of a view and the maximum brightness of the display as seen from the same angle. The visibility of each view is a function of the observation angle. The function gives the visibility of a given view from observation angle θ. The model uses the assumption that the function is the same for all views, with the peak visibility of each view occurring at a different observation angle. In Fig. 9, k_v is used to denote the angular offset in the visibility function for v view n_v. More information on modeling of autostereoscopic displays as image processing channels can be found in [39, 40, 42, 43].

3 Visual Quality of Stereoscopic Displays

The visual quality of a 3D display is determined by its ability to visualize a 3D scene with little or no visible distortions. Most often, the display is used in a so-called *no-reference* setup, i.e. the observer is presented with a (possibly) distorted display replica of a scene but is not presented with the original scene itself; thus a full comparison between the reference and the replica is not possible. Instead, the visual quality is judged on the basis of the presence of recognizable distortions (e.g. artifacts) and the subjective level of annoyance they cause. In this chapter the distortions are categorized according to their origin and they are separated into three large groups:

- *Viewpoint-related* distortions: their visibility depends on the position of the observer with respect to the 3D display. Examples of such distortions are ghosting (due to angular crosstalk), pseudoscopy (due to bad observer position) and accommodation-convergence rivalry. Angular-dependent artifacts are common in autostereoscopic displays since the image generated by such displays is a function of the observation angle. However, SDPG displays are also affected because the performance of the polarization filter depends on the angle. Finally, accommodation-related artifacts affect all 3D displays which do not re-create focal depth cues [44].
- *Multiplexing-related* distortions: these are caused by the process of combining multiple images for presenting onto one display. Sub-optimal channel separation results in some minimal crosstalk regardless of the observation position. Minimal crosstalk is present in both temporally and spatially multiplexed 3D displays. Incorrectly prepared images for spatially-multiplexed displays could exhibit Moiré artifacts due to aliasing. Visible gaps between the sub-pixels or non-rectangular pixel shape manifests itself as masking artifacts (also known as fixed-pattern noise) [45].
- *Content-related* distortions: these are caused in the process of content preparation. It is possible that parts of the stereoscopic image are not fuseable by the

HVS. There are two reasons for this; one is that the disparity is too large and the other is that regions of the scene are close to the frame and are present in one channel only. If the observer tries to focus on such an area, he or she experiences *diplopia*. If that happens for objects with apparent position in front of the screen it is perceived as the *frame violation* artifact that is more annoying than diplopic objects behind the screen [46].

3.1 Visibility of Image Distortions

In this chapter we focus on artifacts which affect stereoscopic perception. However, due to the layered nature of the HVS, stereoscopic artifacts might be induced by monoscopic distortions, for example *blockiness* is a monoscopic artifact visible by a single eye, but can distort display disparity and destroy a binocular depth cue. More information on artifacts in 3D scenes and their taxonomy can be found in [4, 39, 47–51].

3.1.1 Viewpoint-Related Distortions

If two views are simultaneously visible by the same eye the effect is regarded as crosstalk between the views. If an object of the scene is meant to have apparent depth, its representations in each channel have horizontal disparity. The combination of crosstalk and disparity creates a horizontally-shifted, semi-visible replica of the object. The combination of double contours and transparency is interpreted by the HVS as *ghost images*, or ghosting [47]. An example for ghost images is shown in Fig. 10a. If the amount of crosstalk is different for each color channel, the shifted replicas have different colors, as shown in Fig. 10b. This effect is referred to as *color bleeding*. In autostereoscopic displays the visibility of a view is a function of the observation angle, as shown in Fig. 10c. The position where one view has maximum visibility, and the other is maximally suppressed is known as the *sweet spot* of that view. The observation zones of the two views are separated by a zone where neither of the views is predominantly visible. That zone is also known as the *stereo-edge*. For autostereoscopic displays, visibility of the ghosting artifacts is proportional to the crosstalk and has its minimum in the sweet spots and its maximum in the stereo edge. Subjective visual quality experiments described in Kooi [52] and Pastoor [4] suggest that inter-channel crosstalk of 20% is the maximum acceptable in stereoscopic image.

Another viewpoint-related distortion is the so-called *accommodation-convergence (A/C) rivalry*. On a stereoscopic display the distance to the convergence point can be different from the focal distance, as shown in Fig. 11a. This difference is known as *accommodation-convergence mismatch.* The accommodation-convergence reflex drives the eyes to focus at a wrong distance, which causes the objects with pronounced apparent depth to be perceived out-of-focus. A large discrepancy

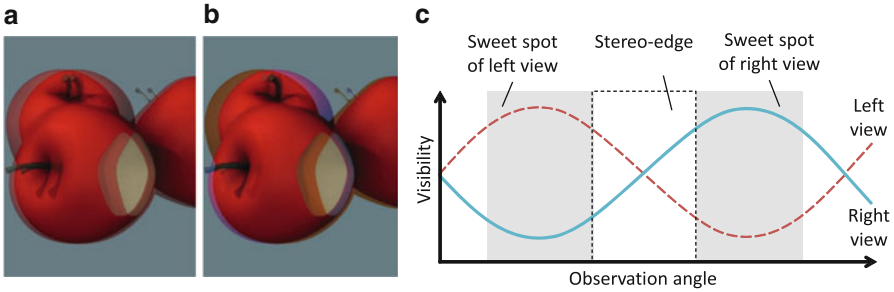


Fig. 10 Ghosting artifacts: (**a**) color-balanced ghosting, (**b**) color bleeding and (**c**) crosstalk as function of the observation angle in autostereoscopic displays

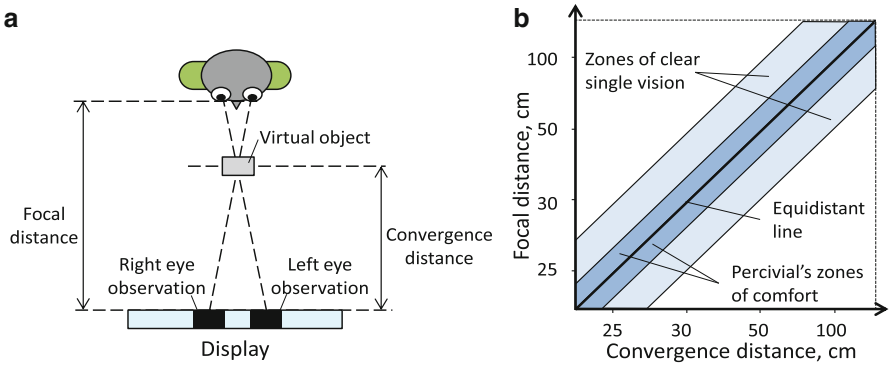


Fig. 11 Accommodation-convergence rivalry: (**a**) focal and convergence distance mismatch and (**b**) zones of clear single vision and Percival's zones of comfort (adapted from [44])

between the focal and convergence distance prevents the eyes from converging, causing *diplopia*. Stereoscopic fusion is possible only for some combinations between focal distance and convergence distance. The set of focal and convergence distances which allow fusion define so-called *zones of clear single vision*, as seen in Fig. 11b [44]. Inside the zones of clear single vision resides a narrower area, known as *Percival's zone of comfort*, where the difference between the apparent and actually focal distance is less than 0.5 diopters. Within the Percival's zone of comfort A/C rivalry is negligible [2, 44].

Pseudoscopy (*reverse stereo*) is the situation in which the eyes see the opposite views; i.e. the left eye sees the right view, and vice versa. For example, the left observer in Fig. 12a sees proper stereo image, while the observer in the right experiences pseudoscopy. In a pseudoscopic image the binocular depth cues contradict the pictorial ones, which results in perceptually disturbing image [47]. Another factor which narrows the size of the sweet spots is the stereo-edge. Between the stereoscopic and pseudoscopic areas there are zones with high crosstalk where the 3D effect is not visible, as marked with "X" in Fig. 12b.

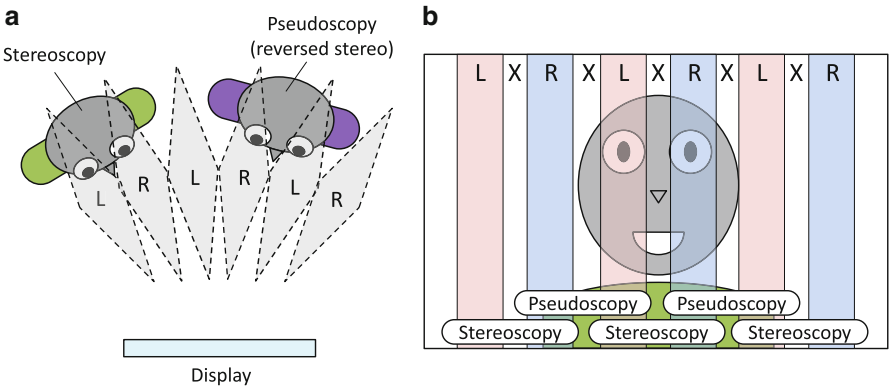


Fig. 12 Pseudoscopy: (**a**) stereoscopic and pseudoscopic observation zones, view from the top and (**b**) observation zones which yield clear stereoscopic image

In addition, some artifacts are most obvious for moving observer; for example the Moiré-like pattern seen on an autostereoscopic display exhibiting *picket fence effect*, or *banding* [35, 47]. Unnatural representation of image parallax causes *shear distortion* in dual-view displays, and *image flipping* in multiview ones [47]. More information about viewpoint-related distortions is available in [4].

3.1.2 Distortions Related to Spatial View Multiplexing

In spatially-multiplexed displays the optical filter introduces selective masking over the sub-pixels of the display, thus separating them into different visual channels. This masking can be modeled as a sub-sampling on a non-orthogonal grid. Without pre-filtering this process creates aliasing artifacts which are perceived as Moiré artifacts.

In multiview displays Moiré artifacts are visible in all types of scene, but are especially pronounced in 2D content as in 3D images aliasing is somewhat masked by more severe artifacts such as ghosting [53]. A visual example of Moiré artifacts is shown in Fig. 13. In Fig. 13a one can see a test image which contains various image details susceptible to aliasing. Knowing which sub-pixels are going to be masked by the optical layer one can simulate the output in multiview 3D displays. Such simulation is presented in Fig. 13b. As one can see, Moiré artifacts are present. Finally, Fig. 13c shows an actual photograph of a 3D display showing the test image from Fig. 13a. The display has a light diffusing layer which slightly blurs the image [36], with the aim to decrease the visibility of Moiré artifacts.

In many autostereoscopic displays, even at the sweet spot of one view, the contours of one or more other views are still visible. The crosstalk level at the best observation position is known as *minimal crosstalk.* The effect of the minimal crosstalk is especially pronounced in multiview displays where the visibility zones

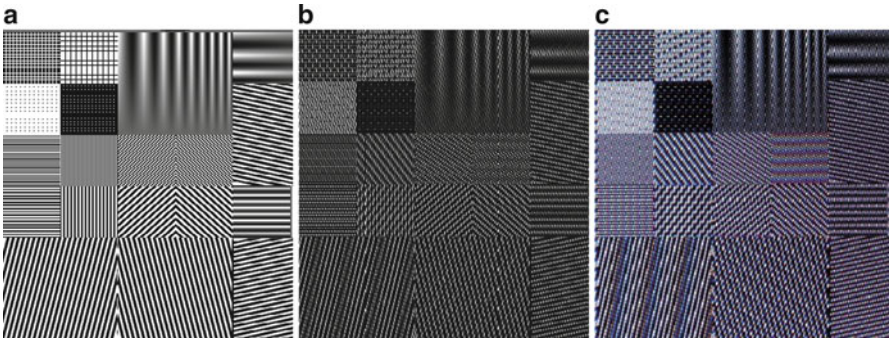


Fig. 13 Moiré artifacts, caused by irregular sub-sampling: (**a**) test image, (**b**) simulated effect of the optical layers and (**c**) actual photograph of a multiview display, showing the test image in (**a**).

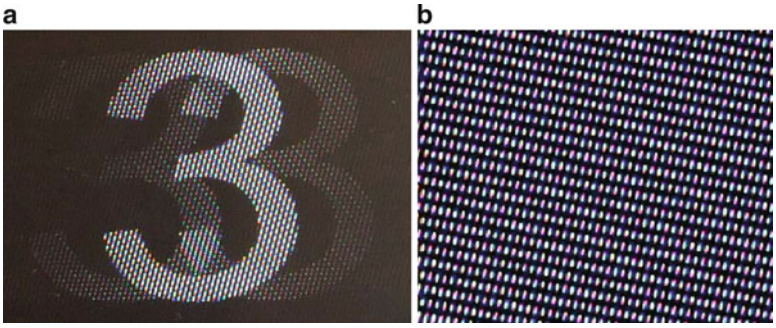


Fig. 14 Photographs of displays with spatial multiplexing, showing typical distortions: (**a**) multiple ghosts of an image and (**b**) imaging, or fixed-pattern noise

of different views are interspersed, and from a given angle multiple views are simultaneously visible [27, 38, 54]. An example image, exhibiting multiple ghosting artifacts is shown in Fig. 14a. The presence of ghosting artifacts degrades the quality of a 2D image but is especially damaging for a stereoscopic image. The presence of repeated edges in horizontal direction introduces ambiguity in binocular disparity and can completely destroy the binocular depth cues [4, 52, 55].

In displays with a parallax barrier, the barrier creates visible gaps between the pixels as seen in Fig. 14b. These gaps are seen as masking artifacts, similar to the fixed-pattern noise exhibited by some digital projectors [45]. The perceptibility of masking is limited by physiological factors such as the optical properties of the eye, the density of photoreceptors and the contrast sensitivity function [56]. However, even if separate elements of the mask are visible, the brain has a limited cognitive ability to reconstruct the underlying shape. That ability is known as the *visual Gestalt principle* [1] and the interdependent visibility of patterns with different properties is modeled as *pattern masking* [56]. More information about distortions, related to spatial view multiplexing can be found in [40, 47, 49, 50, 53, 54, 57, 58].

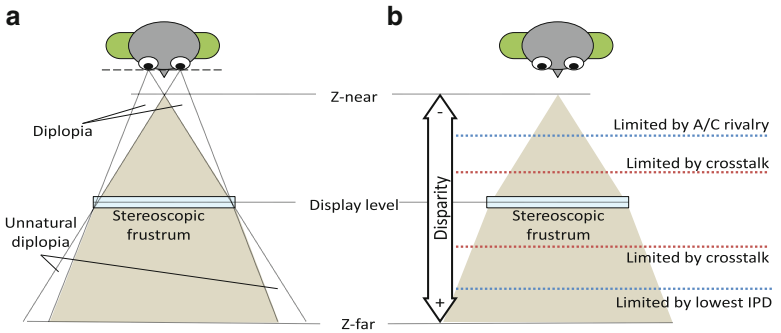


Fig. 15 Disparity range of a comfortably perceived content: (**a**) stereoscopic frustum and (**b**) factors limiting the comfortable disparity range

3.1.3 Content-Related Distortions

For a scene on a 3D display there is a limited space where an object should appear in order that the object is visible in both eyes. This space is known as the *stereoscopic frustum* and is defined by the positions of the eyes and the size of the display, as shown in Fig. 15a. The size of the frustum defines the maximum absolute disparity for objects as a function of their position on the display.

Inside of the frustum there is a limited amount of disparity values that can be present in stereoscopic content in order for that content to be comfortably observed on a given stereoscopic display. In this chapter this range is called the *comfort disparity range* and an example for disparity limits is shown in Fig. 15b. One limiting factor to the comfort disparity range is the A/C rivalry discussed earlier. Another limitation comes from the process of eye convergence; the inward and outward motion of the eyes is limited. Eyes can converge at distances ranging from about 5 cm in front of the head, to infinity. The eye muscles do not allow the eyes to look in divergent directions. The maximum disparity that can be perceived is limited by the observer's IPD. A larger disparity causes *divergent parallax* which is a disturbing, or potentially painful experience [47]. This limitation is somewhat less pronounced in mobile 3D displays as the mean IPD of 65 mm corresponds to substantial part of the display width and the limits imposed by A/C rivalry occur for disparities much lower disparities than the IPD.

The combined influence of A/C rivalry and divergent parallax determines the *objective comfort disparity range*. The subjective experience of a content with excessive disparity is known as *hyperstereopsis* [2] and is considered to be a very disturbing artifact, possibly outweighing all other visual artifacts in 3D content [4, 52]. However, the objectively calculated comfort disparity range does not coincide well with the subjective experience. Apparently, there are many additional factors that influence the comfort disparity range of a stereoscopic display, for example; minimal crosstalk, optical quality, brightness and local contrast of the visualized content. There is another, *subjective comfort disparity range*, usually narrower than

the objective one, which represents the subjective experience of the user and his or her acceptance of 3D content with given disparity. More information about content-related distortions can be found in [52, 55, 57, 58].

3.2 Visually Important Properties of Stereoscopic Displays

The design of a stereoscopic display is a trade-off between observation convenience and visual quality. There are number of previous works that deal with estimation of the optical quality of display and they include theoretical considerations about the interleaving map [40, 42], trough measuring of the optical parameters of the display [54, 59, 60] and subjective tests with different multiview displays [4, 52, 55]. However, evaluating the quality of a 3D display solely based on its optical parameters has two main disadvantages; (1) some optical parameters are directly related to the perceived quality and (2) visibility of 3D artifacts depends on other factors as well, for example scene content and observation conditions.

Gaining knowledge of 3D display parameters serves two goals. One goal is to allow the consumer or content producer to compare the visual quality of two displays, or judge if a given 3D content is suitable for a certain display. The other goal is to use signal processing techniques to mitigate the artifacts in a given 3D display, thus optimizing the visual quality of the output. This section aims to identify the display characteristics, significant from image processing point of view, and relate them to visual quality.

3.2.1 Position and Size of the Sweet Spots

In stereoscopic displays the *optimal observation region* is the observation position where the stereoscopic image is perceived with sufficient quality. In passive autostereoscopic displays these regions are small and distinct areas also known as *sweet spots*. However, optimal observation regions also exist in glasses-enabled 3D displays; for example, the crosstalk in a SDPG depends on the observer's elevation. According to [4, 52], 20% of crosstalk is the limit of crosstalk acceptance for 3D displays. In this chapter we define the sweet spot as an observation position where each eye perceives the proper view and the crosstalk between the views is less than 20%.

Since the display is flat, from a given observation position different parts of the screen surface are seen from a slightly different observation angle, as shown in Fig. 16a. The viewing zone of a view is formed by the union of the visibility zones of each pixel that belongs to that view and has a characteristic diamond-like shape, sometimes referred to as *viewing diamond* [54]. For stereoscopy to be possible each eye needs to be in the corresponding sweet spot, as seen in Fig. 16b. This requirement imposes a limit on the range of observation distances suitable for a given display. The size of the sweet spots can be derived from the angular visibility

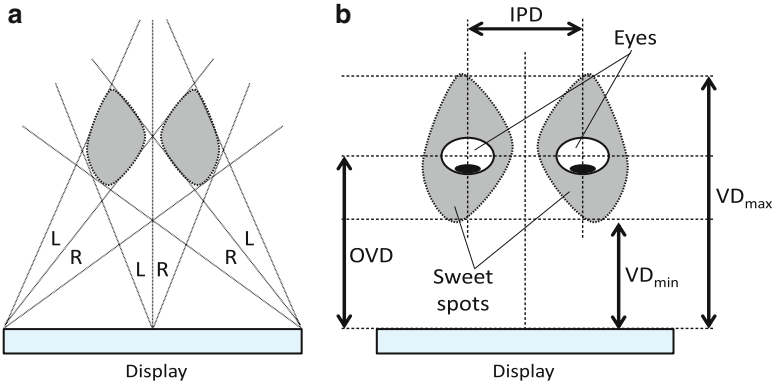


Fig. 16 Sweet spots of an autostereoscopic display: (**a**) left and right sweet spots and (**b**) optimal, minimal and maximal observation distances

function, or directly measured using a pair of cameras separated at the IPD. For a given IPD there would be minimal and maximal distance at which both eyes on the observer appear inside the corresponding sweet spot. These viewing distances are marked in Fig. 16b as VD_{max} and VD_{min}. Also, for a given IPD there would be an optimal observation distance at which there is an optimal optical separation and lower crosstalk visible across the whole surface of the display. The optimal viewing distances is labeled as OVD in Fig. 16b. Usually OVD, VD_{max} and VD_{min} are calculated using the mean IPD of 65 mm.

Naturally, the size and position of the sweet spots is related to the perceived quality. As discussed in [61], a 3D display with a few, larger sweet spots is considered easier to use than another display that has many sweet spots of smaller size. More information about measuring and modelling of 3D display sweet spots can be found in [35, 39, 54, 57, 59].

3.2.2 Interdigitation Map

The map indicating the relation between the position of a sub-pixel and the view it belongs to is known as *interdigitation map*. Since both the TFT–LCD and the optical filter have repetitive structure, the interdigitation map is built from a smaller, repetitive *interdigitation pattern*. The pattern is spatially independent; angular visibility of a sub-pixel depends on its position in respect to the pattern, but not on its absolute position in respect to the display. The interdigitation map ranges from simple ones for dual-view displays (see Fig. 17) to complex ones for multiview displays (see Fig. 18). Most SDPGs have row-interleaved topology such as the one shown in Fig. 17a, as such topology ensures higher horizontal resolution. Autostereoscopic displays have column-interleaved topology since they rely on parallax-based light redirection, and views should be separated in horizontal

a

	Pixel 1			Pixel 2		
	R	G	B	R	G	B
1	L	L	L	L	L	L
2	R	R	R	R	R	R
3	L	L	L	L	L	L
4	R	R	R	R	R	R

b

	Pixel 1			Pixel 2		
	R	G	B	R	G	B
1	L	L	L	R	R	R
2	L	L	L	R	R	R
3	L	L	L	R	R	R
4	L	L	L	R	R	R

c

	Pixel 1			Pixel 2		
	R	G	B	R	G	B
1	L	R	L	R	L	L
2	L	R	L	R	L	R
3	L	R	L	R	L	L
4	L	R	L	R	L	R

Fig. 17 Interdigitation maps of dual-view autostereoscopic displays: (**a**) row-interleaved, (**b**) column-interleaved at pixel level and (**c**) column interleaved at sub-pixel level

a

$$A=\begin{bmatrix}2 & 5\\ 4 & 7\\ 6 & 9\end{bmatrix};\ B=\begin{bmatrix}8 & 11\\ 10 & 13\\ 12 & 15\end{bmatrix};$$

$$C=\begin{bmatrix}14 & 17\\ 16 & 19\\ 18 & 21\end{bmatrix};\ D=\begin{bmatrix}20 & 23\\ 22 & 1\\ 24 & 3\end{bmatrix}.$$

b

	Pixel 1			Pixel 2			Pixel 3			Pixel 4			Pixel 5			Pixel 6			Pixel 7			Pixel 8		
	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B
1	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
2																								
3																								
4	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A
5																								
6																								
7	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B
8																								
9																								
10	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C
11																								
12																								

Fig. 18 Interdigitation pattern of a multiview display: (**a**) interdigitation patches, describing correspondence between view number and sub-pixel position in the patch and (**b**) position of the patches in the interdigitation pattern

plane. Pixel-based column interleaving as shown in Fig. 17b results in imbalanced color separation and produces color bleeding artifacts. Sub-pixel-based interleaving (shown in Fig. 17c) does not suffer from color bleeding. Note that autostereoscopic displays can have row-based interleaving as well, provided that the TFT–LCD matrix is rotated at 90° so its pixel columns appear horizontal [62].

Multiview displays have slanted interdigitation topology where sub-pixels from one view appear along a slanted (in respect to the TFT) line. In order to prevent color bleeding the horizontal size of the interdigitation pattern is not divisible by 3, e.g. neighboring sub-pixels from the same view and on the same row have different color, as can be seen in Fig. 18. As a result, pixels from one view appear on a non-rectangular grid. In order to design proper sub-sampling filter for that grid one needs to know the precise interleaving topology of the display [35, 40, 43].

The interdigitation pattern is given in a compressed form. The patches in Fig. 18a describe the correspondence between sub-pixels and the view numbers, for example the top-left sub-pixel in patch “a” belongs to view 2. The map shown in Fig. 18b gives the position of these patches in the interdigitation pattern.

Ideally, each view should be seen with full brightness from its visibility zones (marked with “V” on Fig. 19a) and be invisible from anywhere else (as marked with “N” in the same figure). A group of sub-pixels with similar angular visibility will have higher N/V ratio than a group of sub-pixels with varying optimal observation

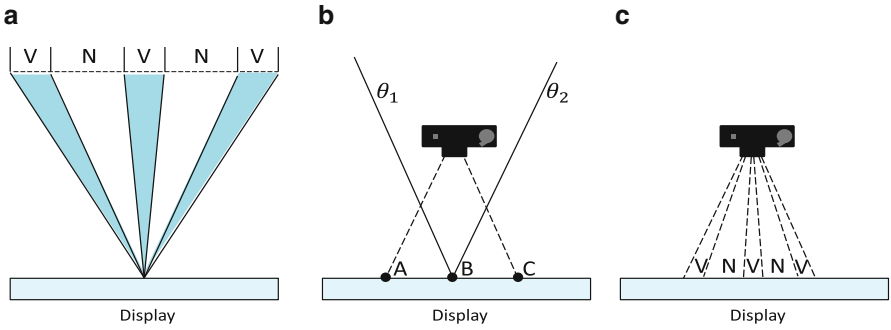


Fig. 19 Finding the interdigitation topology: (**a**) visibility zones of a view, (**b**) angular visibility of points in a close shot and (**c**) N/V ratio as seen in a close shot

vectors. This allows one to find the sub-pixels which belong to a particular view. This can be done by turning a group of sub-pixels on and observing the angular visibility of the resulting image. Instead of measuring the visibility of a view from multiple angles, one can photograph the display from a distance shorter than the optimal observation one, as shown in Fig. 19b. Following the assumption for spatial independence of angular visibility, the visibility points along the horizontal axis would correspond to the visibility of one point as seen from different angles. As exemplified in Fig. 19b, point "A" as seen from the camera should be the same as the visibility of point B as seen from observation angle θ_2, and point "C" as seen from the camera should be the same as the visibility of point B as seen from observation angle θ_1. In the photograph the ratio between visible and invisible parts is proportional to the N/V ratio of the pixel group under test, as shown in Fig. 19c. The group of sub-pixels with the highest N/V ratio belongs to the same view. More details about the procedure for finding the interdigitation topology of a 3D display can be found in [63].

3.2.3 Angular Visibility

In this subsection we present a simple, yet efficient way of measuring the angular visibility of a multi-view display using an off-the-shelf camera. The angular visibility function of each display element allows one to predict the position of the sweet spots and the crosstalk for different observation positions. Measuring the brightness of a single pixel by photographing the display would be a tedious and noise-prone task. Instead, one could measure the mean brightness of a view and assign it to each pixel of that view since sub-pixels in one view are supposed to have the same angular visibility. Another problem arises when measuring at different angles; if camera position is inaccurate the angular visibility curve would be sampled at irregular intervals. This can be solved by measuring the visibility of

each view at some selected points and searching for single function that gives the best fit for all measurements, regardless of the angle.

The first step of the measurement technique is to prepare two groups of test images. The first group consists of so-called *single view* images where only the sub-pixels from one view are lit. These images are used for measuring the angular visibility. The second group contains test images where all pixels are set to different levels of grey in order to linearize the camera response function [64]. In the second step, each test image is shown on the test display and is photographed from a number of observation positions. The observations positions are selected on a line parallel to the display surface and at the optimal viewing distance. If the measurement point is displaced from the center of a visibility zone, the visibility function gets sampled with an offset and the maximum value of that function falls in between two samples. However, judging by measurement results in other works [54, 59, 65], one can assume that the visibility function for all observation points can be closely approximated by the same function, which has its peak occurring in the optimal observation spot for the corresponding view. In the third step, based on this assumption, one can search for single function that closely fits measurements for all positions regardless of possible offset.

More details about measuring the angular visibility function of a 3D display can be found in [35, 54, 59, 60, 63].

3.2.4 Display Passband

Spatially-multiplexed 3D displays suffer from masking distortions and fixed-pattern noise caused by visible gaps between the pixels and/or by apparent non-rectangular shape of a pixel. The visibility of such distortions depends on interaction between the spectrum of the visualized content and the display's transfer function. This interaction can be conveniently expressed in the frequency domain. Therefore, in order to assess the visibility of masking, one needs to study the performance of the display in frequency domain through a quantity called a *display passband*. In this subsection we present a simple yet efficient six-step methodology to measure the display passband. The approach is shown in Fig. 20.

The first step is to prepare a number of *test signals* which contain a 2D sinusoidal pattern with varying horizontal and vertical frequency components, as the ones shown in Fig. 21a, c. Then, out of each test signal a number of *test images,* each one with different apparent depth, are prepared. This is done by mapping the same signal to each view of the display, adding different amount of disparity to each view and interleaving all views in a test image. The third step involves automated visualization of all test images on the display and making a snapshot of each one with a high-resolution camera. The output of that step is a collection of *test shots* of all test images, similar to the ones shown in Fig. 21b, d. In the next step the *spectrum* of each test shot is analyzed in order to determine the amplitude ratio between the original frequency component in the test signal and the most noticeable distortion frequency component introduced by the display.

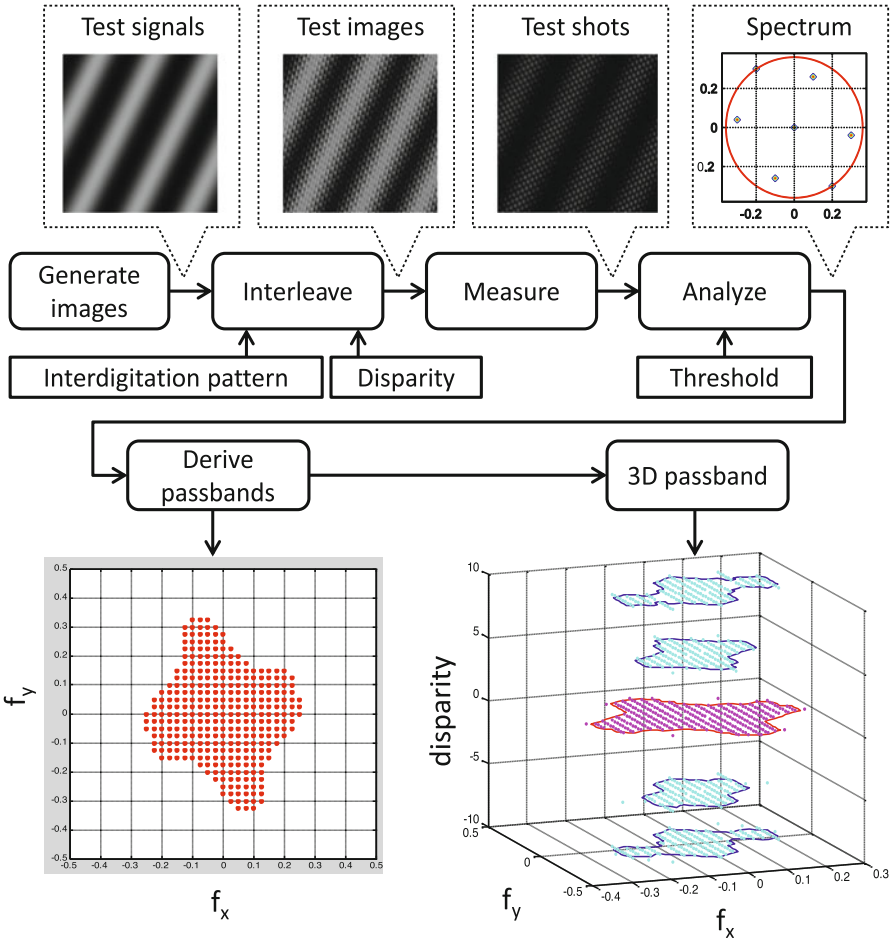


Fig. 20 Block diagram of the methodology for deriving of the display passband

The distortion frequency is selected as the largest peak in the spectrum, which is positioned closer to the center than the original frequency component of the input signal. Based on the distortion-to-signal ratio (ration between magnitudes of the distortion and signal frequency), the so-called display passband is derived. Frequency components of the test image with ratio smaller than a threshold are marked as being inside of the passband, and otherwise as being outside. For example, the test pair shown in Fig. 21a (test), b (observation) belong to the display passband as the frequency components in the text image are still dominant in the observation. As the HVS can reconstruct missing elements of a structure, the horizontal bars in Fig. 21a are still visible in Fig. 21b. The test pair shown in Fig. 21c, d does not belong to the passband since the dominant frequency in the test image is masked by the distortion. The passband is scanned by analyzing the

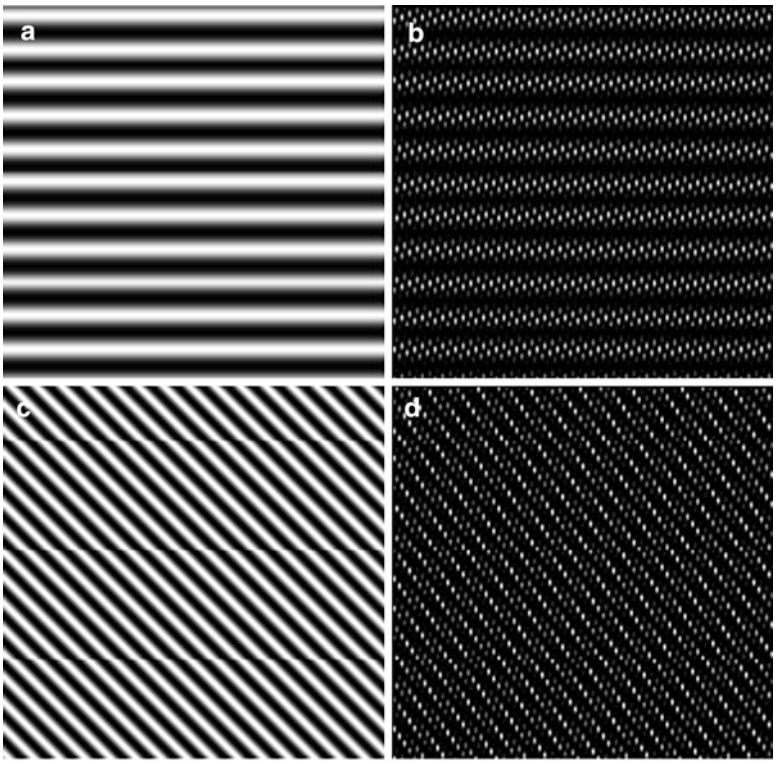


Fig. 21 Deriving display passband: (**a**) test image with horizontal frequency component, (**b**) observation of the first test image, where the intended frequency is still dominant, (**c**) second test image and (**d**) observation of the second test image, where the intended frequency is masked by distortions

distortion-to-signal ratio for multiple test images for various frequencies (sampled on a dense grid). In step 5 all input frequencies which passed the threshold are combined into display passband area, as shown in Fig. 20, bottom-left. The passband area represents the ability of the display to faithfully reproduce image signals with spatial frequencies within the area. Finally, in the sixth step, all passbands measured for different disparities are collected into a 3D passband area, as shown in Fig. 20, bottom-right. The shape and the size of the 3D passbands enable quality comparison between 3D displays. A display with a larger and more uniform passband would be of higher visual quality as it can faithfully represent larger range of image details. Additionally, by knowing the frequency characteristics of a 3D scene, content producers can judge if the scene would "fit" the passband of a given display, resulting in a faithful representation. More details about deriving the passband of a 3D display can be found in [49, 63].

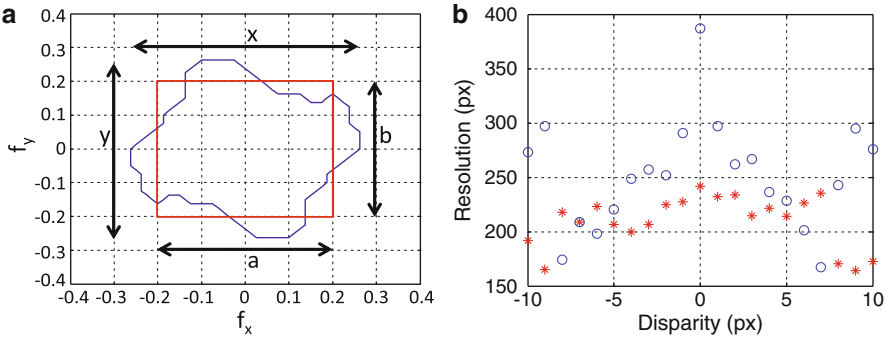


Fig. 22 Equivalent resolution of a 3D display: (**a**) fitting rectangle to the passband and (**b**) equivalent resolution in horizontal (*circle*) and vertical (*star*) direction as a function of disparity

3.2.5 Equivalent Perceptual Resolution

Although the display passband allows a quality comparison between displays, it is not straightforward to use it for judging the quality of a single display. Since most display users have an intuitive idea about the image quality of a display with a given resolution it is beneficial to convert the 3D passband into a "corresponding" 2D display resolution. This can be done by approximating the passband for each disparity with a rectangular shape. The main idea is to have a rectangle centered at origin that will have the same area (in size) as the original passband, trying to overlap as many passband points as possible. Another requirement is to keep, at the same time, the aspect ratio between maximum values in horizontal and vertical direction. With these two constrains (area and aspect ratio) one can find the rectangle which is the "best fit" to a given passband area, as shown in Fig. 22a. More details on approximating the passband with rectangle can be found in [49].

In order to represent this figure in a more understandable way, one can convert rectangular passband sizes to equivalent resolution in pixels. This is done by multiplying the passband width (height) with the overall resolution in horizontal (vertical) direction. An example for equivalent resolution as a function of the disparity as derived for a 24-view 3D display can be seen in Fig. 22b. Notably the function is not monotonic but has local maximums for some disparities. Knowing the equivalence resolution of a 3D display can help content producers to rearrange placement of objects in a 3D scene so that each object is seen with optimal quality.

3.2.6 Comfortable Disparity Range

There are a number of parameters which determine the maximum disparity range which can be comfortably observed on a 3D display. Some of them such as divergent parallax, A/C rivalry and frame violation can be calculated provided that one knows the display resolution, pixel density, observation distance and the IPD of

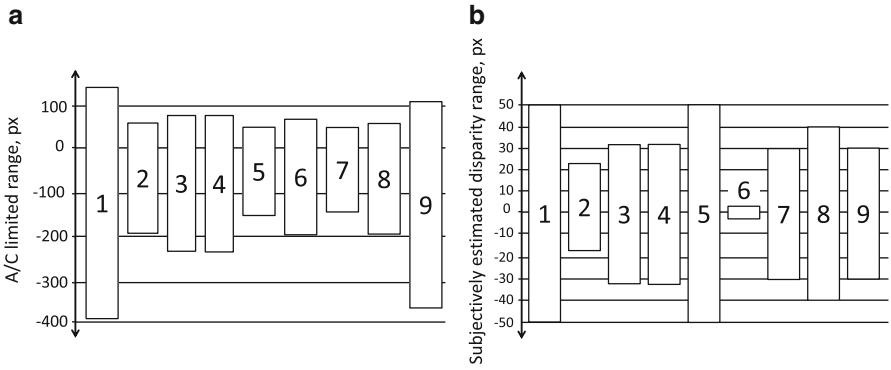


Fig. 23 Comfortable disparity range of various 3D displays: (**a**) calculated using objective parameters and (**b**) derived from a subjective test

the observer. However, other (and less studied) parameters are probably involved as well, for example; subjectively perceived contrast, screen reflection index, room illumination, etc. The unambiguous way to determine the comfortable disparity range of a 3D display is to perform subjective tests where the acceptance of 3D content is rated. Naturally, the main variable in the experiment is disparity range. Since local contrast of the content greatly influences the perceptibility of ghosting artifacts [4, 47, 52], the content under study should contain scenes with various levels of contrast. As contrast perception is frequency dependent [66], acceptance of 3D content is possibly affected by the frequency characteristics of the image.

An example of a subjective experiment is presented below. A group of 10 observers was asked to rate the acceptance of a number of test images. The images contain two patterns; a text pattern and a natural scene pattern. With each pattern a number of images with varying local contrast are created. The contrast is altered by changing the brightness of the patch and that of the background. Finally, each test image is used to generate a number of stereoscopic pairs with varying disparity. Observers were asked to rate each stereoscopic pair. The test was repeated for nine different 3D displays.

The comfort disparity range for each display was calculated using objective parameters, such as minimal IPD and Percival's zone of comfort. The group of ranges is shown in Fig. 23a. From the subjective test, the subjective disparity range was derived for each display, as shown in Fig. 23b. The displays included in the experiment are as follows; (1) large SGPG 3D TV set, (2) mobile 3D display prototype with horizontally double-density pixel (HDDP) pixel arrangement, (3) commercial mobile display with switchable parallax barrier and operating in landscape mode, (4) same as the previous model but operating in portrait mode, (5) 3D photo-frame with parallax barrier, (6) commercial 3D camera with stereoscopic viewfinder, (7) laptop with autostereoscopic 3D display, (8) Laptop with SDPG 3D display and (9) prototype of pocket media player with autostereoscopic 3D display.

From the figures, one can see that the subjective comfort disparity range is 4–5 times smaller than the objectively calculated one. Apparently, the influence of display properties influences the range more than the viewpoint-related parameters. More information about comparing various parameters of 3D display can be found in [30, 38, 57].

4 Visual Optimization by Signal Processing

Signal processing techniques can be used for improving the visual quality of 3D displays in three ways. If a distortion introduced by the display can be described as an invertible function, one can pre-process (pre-distort) the image using the inverse function. In such case, the changes caused by pre-processing would cancel display distortions, resulting in a clean signal representation without artifacts. Such process is known as *pre-compensation* and can be used to improve some cases of pseudoscopy, hyperstereopsis and ghosting. In the case of distortions which cannot be pre-compensated, a signal processing algorithm can decrease their visibility, helping mitigate the perceived annoyance of artifacts, thus improving the quality. Artifact mitigation algorithms are possible for imaging, aliasing and cases of pronounced crosstalk. Finally, the visibility of some artifacts does not depend purely on the content but also on observer position, motion and head orientation. Such cases need real-time algorithms which actively track the observer and process the visual signal accordingly.

A list of artifact mitigation techniques is given in Table 1. In order to mitigate distortions caused by observation angle one needs to know the position of observer in respect to the display. Most often this is done by using camera-based tracking and face- or eye-tracking algorithms. Once the observation position is known the image can be optimized for the calculated angle and distance. Although user-tracking displays that can work with up to four observers exist [33], algorithms for viewpoint optimization usually work for one observer only. Ghosting artifacts can be either pre-compensated or mitigated. For dual-view displays, where crosstalk levels are low, pre-compensation is possible but limits the dynamic range of the display [67]. Crosstalk pre-compensation is possible both for time-sequential and spatially-multiplexed dual-view 3D displays. A similar approach can be used for a multiview display if a single observer is tracked. However, the possibility of multiple observers and the pronounced crosstalk between neighboring views make crosstalk mitigation the preferred approach for multiview 3D displays. Such algorithms aim to reduce the visibility of ghost images by filtering horizontal high-frequency components of the image but at the expense of losing image details.

The range of artifacts which are caused by the optical separation layer of a multiview display can be mitigated by antialiasing filters [40, 43, 68], or by deriving the passband of the display, and prepare a filter which removes image data with frequency components outside of the passband [69]. Such a filter can be implemented as a single 2-D filter [70], or as a bank of 2-D filters for various

Table 1 Visual optimization: distortions, artifacts, and mitigation algorithms

		Artifact mitigation algorithm	
Distortion source	Artifact type	Dual-view displays	Multiview displays
Observation angle	Pseudoscopy, ghosting	Pseudoscopy correction	Extended head parallax, extended viewing distance
Crosstalk	Ghosting	Pre-compensation	Crosstalk mitigation
Aliasing	Moiré	Antialiasing filters (1D)	Antialiasing filters (2D)
Interdigitation pattern	Moiré, ghosting, FPN	Pass-band optimization (2D/3D)	
Excessive disparity	Hyperstereopsis	Content repurposing	

disparity levels [69, 71]. If the scene is represented as an epipolar volume, one can implement pass-band optimization as a 3-D filter [68]. Finally, excessive disparity can be compensated by a transformation which alters the disparity range of a scene. Such transformation can be a combination of image rescaling and cropping or, if more processing power is available, a combination of dense depth estimation and image warping algorithms [46].

4.1 View-Point Optimization

In order to adapt the display to the observation position of the user, an artifact mitigation algorithm should detect and track the position of observer's eyes. The eye-tracking should work in real-time because a tracking delay might optimize the image for wrong observation position and introduce visible artifacts. Multiuser observer tracking algorithms have been discussed in [72] (using head-tracking) and [73] (using eye-tracking). In [74] a real-time face and eye-tracking algorithm working on a mobile platform is presented. The implementation allows splitting the processes of face and eye detection between the ARM and digital signal processor (DSP) cores of an OMAP 3430. In order to increase the face detection speed the algorithm searches for a subset of all possible face sizes within the sweet spot of the display and the user is required to stay within the sweet spot of the display. Face detection is performed by a two-stage hybrid algorithm which combines skin detection with feature-based face detection [75] is implemented on the ARM core. If a face is present, eye-detection is performed only in the top half of the detected region. The eye detection is implemented on the DSP core which detects the eyes using a Bayesian classifier working on dual-tree complex wavelet transform (DT-CWT) features [76, 77]. The combination of both algorithms allows precise detection of the position of the eyes in respect to the camera.

4.1.1 Optimization for Observation Angle

Visual optimization for observation angle is solved differently for dual-view and for multiview 3D displays. In dual-view displays the most pronounced viewpoint related distortions are pseudoscopy and ghosting. Ghosting artifacts are seen if either of the observer's eyes appears in the stereo-edge (between visibility zones of two views). Pseudoscopy is seen if both eyes appear in visibility zones of the opposite view. In all other cases both eyes appear in the visibility zone of the same view and a 2D image is perceived. One interesting feature of dual-view autostereoscopic displays is that some models allow switching between the 2D and 3D mode; this allows the display to "fall back" to 2D image and regain display resolution.

An algorithm for observation angle-based optimization for dual-view 3D displays is proposed in [74]. Based on the horizontal coordinate of the pupil, three tracking zones are defined; visibility zone of the left view (marked with "L" on Fig. 12b), visibility zone of the right view (marked with "R" on the same figure), and zone with high crosstalk (marked with "X"). Pseudoscopy is avoided by flipping the left and right channel if the eye is detected to be in the opposite viewing zone. Ghosting artifacts are avoided by turning the parallax barrier off and switching the content to "2D" if either of the observer's eyes appears in an "X" area. The rationale for this rule is that if one eye of the observer perceives excessive crosstalk, stereoscopic perception is not possible and it is preferable that the observer does not see the ghost artifacts either.

In multiview displays the observation zones of neighboring views are interspersed and it is difficult to compensate for ghosting artifacts in real-time. Such displays can provide limited head parallax. However, severe ghosting is visible at the edges of the area where head parallax is experienced. A "semi-active" approach for extending head parallax and removing the ghosting in the stereo-edge is proposed in [61]. It combines the precise light redirection of a multiview display, a single camera and less precise sub-real-time head-tracking. The software part of the system ensures that the observer's head is "surrounded" by a group of properly rendered views. Once the approximate position of the observer's head is found, the precise delivery of different images to the eyes is handled by the (passive) multiview optics. There are a few observation angles where the visibility zones of the first and the last views appear next to each other. A moving observer which crosses one of these boundaries experiences a break of the smooth head parallax [65]. However, one can provide a continuous parallax by replacing the views which are not visible with observations of the same 3D scene from new angles. For example, when the user's head is positioned as seen in Fig. 24a, the active views are from 2 to 5, and views 3 and 4 are seen by the left and right eyes correspondingly. When the user moves to the position shown in Fig. 24b, view 5 shows the 3D scene at the same angles as before and views 6 to 8 are updated to show the scene at a new angle. In reality, the eyes of the user fall into neighboring views and the view update happens well outside of the eye position. The head tracking has only to ensure the head of the observer is approximately at the center of the set of updated views. Unlike the "active" eye-

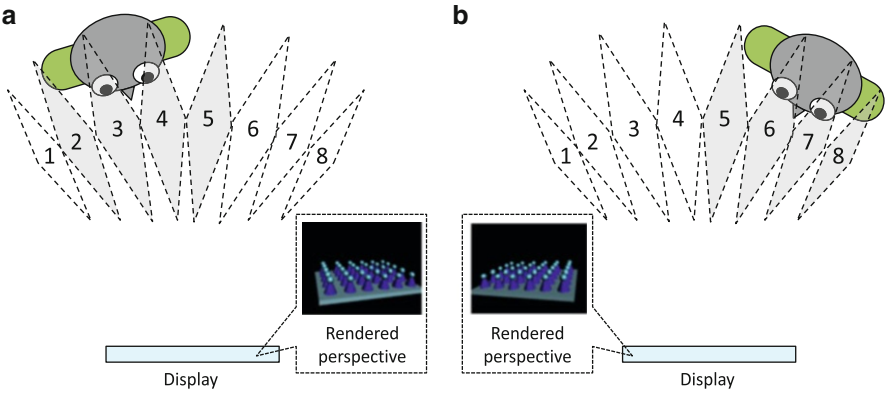


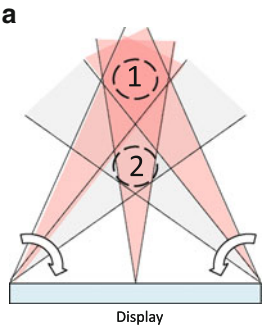
Fig. 24 Selective view updating for continuous parallax: active views and visualized scene perspective for (**a**) one observer position and (**b**) another observer position

tracking approach, estimation of the distance between the observer and the display is not needed as a set of properly rendered views can provide proper parallax to the eyes in a wide range of head positions. Also, real-time performance of the system is not necessarily critical as the user is always "surrounded" by a safe margin of properly rendered views.

4.1.2 Optimization for Viewing Distance

Both dual-view and multiview autostereoscopic displays are designed to be watched at a particular distance. At the optimal viewing distance the intended view is seen across the whole surface of the display, as marked with "1" on Fig. 25a. At a distance closer than the optimal the observer sees different visibility zones at the left and right edges of the display, as marked with "2" on the same figure. If the distance to the observer is known, the content on the display can be re-rendered accordingly. In order to measure the distance to the observer, eye and face tracking is performed by two cameras simultaneously. For more information on the algorithm the reader can refer to [78].

In the case of a multiview display the information is shifted between the views; for example, the image along the right edge of the display intended for the central view (marked with red on the Fig. 25a) can be rendered in the previous view (as shown by the curved arrow). The opposite is done along the left edge. This procedure can be expressed as a re-routing table which optimizes the image for a given observation distance. The re-routing table should be re-calculated for any given distance to the observer. In the case of a multiview display, pixels intended for certain view would be re-routed to other views. An example of a multiview rerouting table is given in Fig. 25b. The surface of the display is separated into sub-sections



b

2	1	0	0	0	0
2	1	0	0	0	-1
2	1	0	0	0	-1
2	1	0	0	-1	-2
1	0	0	0	-1	-2
1	0	0	0	-1	-2

c

F	X	0	0	X	F
F	X	0	0	X	F
F	X	0	0	X	F
F	X	0	0	X	F
F	X	0	0	X	F
F	X	0	0	X	F

Fig. 25 Distance-based content optimization: (**a**) re-routing of views for observation distance, shorter than the optimal, (**b**) example re-routing table for multi-view display and (**c**) example rerouting table for stereoscopic display

and the number in each subsection is an instruction which operation to be performed in the corresponding area of the display.

In the case of stereoscopic display, the re-routing table looks like the one given in Fig. 25c. In this table, "0" means that the pixels in the corresponding area are left unaltered. The pixels in the "F" areas should be "flipped", effectively swapping the pixels intended for the left and right view. The areas marked with "X" would be perceived with excessive crosstalk because for these areas the observer appears between the viewing zones of the left and right views. In the "X" areas a monoscopic image should be projected by copying all pixels from one view to the other.

4.1.3 Optimization for Observation Pose

Some dual-view autostereoscopic 3D displays with a parallax barrier have the ability to switch between horizontal 3D and vertical 3D modes [34]. For such displays a visual optimization algorithm can select the 3D mode and scene orientation based on the orientation of the observer's eyes, as illustrated in Fig. 26a. If the face of the observer is not in the horizontal or vertical direction in respect to the display the 3D effect is not possible and thus the system switches the display into 2D mode. An eye-tracking algorithm for selecting scene orientation is proposed in [74]. The block diagram of that algorithm is shown in Fig. 26b. First, face detection is attempted four times, each time rotating the input image by 90°. If detection fails, the presumption is that either the face of the observer is too far from the display or it is at a wrong angle. In both cases 3D perception is not possible and the system switches the display into 2D mode. If face detection is successful its direction is stored and eye tracking is performed according to the direction. The position of the eyes is matched against the map of observation zones of each view (see also Fig. 12b). The map in use is selected to match the direction of the face. If both eyes are found in the corresponding regions the system switches into 3D mode. If both eyes appear in the regions of the opposite view the system flips the channels and

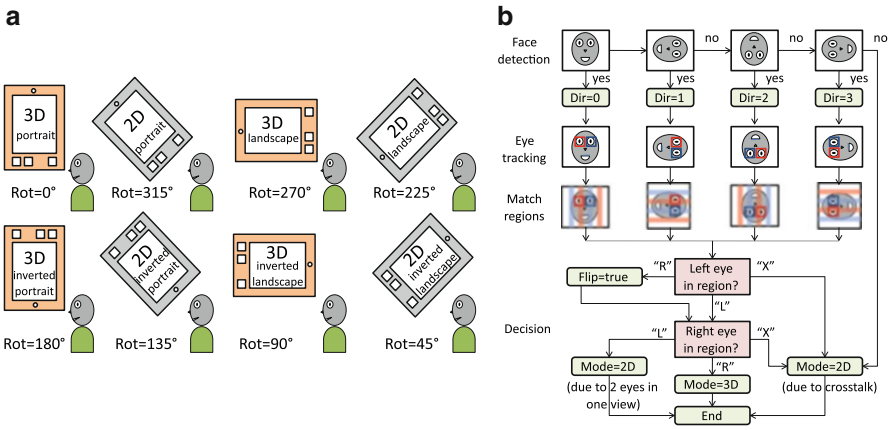


Fig. 26 Selection of display mode and scene orientation, according to the orientation of the eyes of the observer: (**a**) mode and orientation according to rotation angle and (**b**) block diagram of the algorithm

activates the 3D mode. If both eyes fall into the observation zone of the same view, or at least one eye falls in an inter-zone crosstalk area, the system switches into 2D mode.

4.2 Optimization of Display Passband

When visualizing images on spatially-multiplexed displays there are two potential sources of distortions; aliasing, due to decimation at sub-pixel level (pixels visible in one view are only a subset of all pixels) and imaging,[1] due to the presence of gaps between pixels. Aliasing can be fully tackled by an anti-aliasing pre-filter. In [53] Jain and Konrad introduced a method for designing 2D non-separable antialiasing filters for an arbitrary sub-sampling pattern. They devised a 2D filter with a passband that spans all frequencies at which the contribution of all alias terms is smaller than the original signal itself. In [43] Moller and Travis used a simplified optical filter model to analyze display bandwidth and derived a spatially-varying 2D filter which requires knowledge of scene per-pixel depth. In [68] Zwicker et al. proposed a low-pass filter to be applied on the sampling grid of the multiview display, expressed in ray-space, which aims at preventing both intra- and inter-perspective aliasing. However, their model does not take into account the directionally dependent aliasing caused by the slanted optical filter.

[1]The effect of the gaps is similar to the one caused by upsampling in the absence of a post-filter. In sampling and interpolation literature the effect is denoted as "imaging" and the filters tackling it are known as anti-imaging filters.

In signal processing, imaging is tackled by an anti-imaging post-filter. In spatially-multiplexed displays, imaging is created by the physical structure of the display; it is therefore impossible to impose a post-filter. However, the visual perception of imaging can be partially mitigated by a pre-filter. Consequently this filter can be merged with the anti-aliasing pre-filter. In order to determine the properties of the required combined (anti-aliasing and "anti-imaging") 2D filter it is necessary to determine the performance of the display in the frequency domain; that is, we have to know which frequency components in the image we can keep (ones that will be properly represented on the screen), and which ones we have to attenuate as potential causes of distortions. A proper design of the filter should result in the best possible representation of images on the display, minimizing aliasing, imaging and ghosting. It is worth mentioning once more that it is impossible to remove imaging artifacts since they are caused by the display's optical layer. However, some of them can be reduced to a level at which they are less disturbing.

A measurement-based method for deriving the frequency response of the display (display passband) is described in Sect. 3.2.4. In the text, the region containing frequencies that are properly represented on the screen is denoted as *passband*, and all other regions as *stopband.* In order to improve the image quality one should design a filter which attenuates frequency components in the stopband. The methods in the following two sections present an example approach for designing such filter.

4.2.1 Passband Approximation with a Non-Separable Filter

The design of non-separable passband-optimizing filter is discussed in [69] and [70]. As a practical example, such a filter is designed for a 24-view 3D display which has a passband as the one in Fig. 20, bottom-left. For that display the shape of the ideal 2D antialiasing filter is as shown in Fig. 27a. In this figure the curve shows the ideal cut-off frequency; that is the passband of the filter should be inside the contour and its stopband everywhere else. For designing a non-separable 2D filter approximating this ideal one, the windowing design technique with the Kaiser window of length 24 has been used [79]. The Kaiser window has been selected as a good candidate due to its relatively narrow transition band and flexible attenuation. The variable parameter of the Kaiser window controlling the stopband attenuation has been set to $\beta = 2.2$. Such selection ensures a stopband attenuation of at least 30 dB that is good enough for the display under consideration. A filter size of 24 by 24 has been chosen as a good compromise between the implementation complexity, transition bandwidth and approximation of the ideal filter.

The design results in the 24 by 24 2D non-separable filter with an impulse response, as shown in Fig. 27b. The corresponding magnitude response (contour) of the designed filter is shown in Fig. 27c. The -6 dB line in Fig. 27c approximates the ideal cut-off frequency. Due to the finite transition bandwidth of the designed filter, even after applying it to the input image, some aliasing errors will occur on the display. However, the aliased frequencies will be attenuated by the filter (either filter transition band or stopband) and as such they will not be visually disturbing.

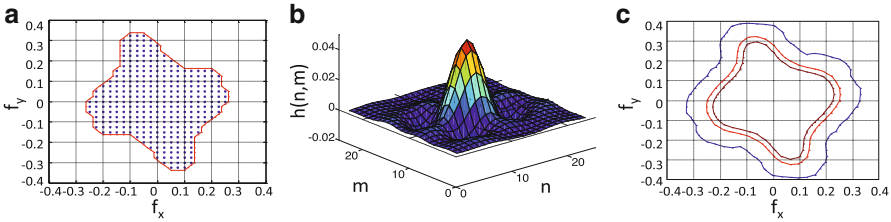


Fig. 27 2D non-separable filter: (**a**) ideal, (**b**) impulse response and (**c**) magnitude response: contour plot for −3 dB (*inner curve*), −6 dB (*middle*) and −30 dB (*outer curve*)

4.2.2 Passband Approximation with a Separable Filter

The computational complexity of a 2D filter is rather high and considerable computational savings can be achieved when it is possible to separate the 2D filter into two 1D filters, one filtering in the horizontal direction and one in the vertical direction. One approach to the design of separable passband-optimizing filter is discussed by Boev et al. [69, 70].

Based on the known interdigitation pattern and angular visibility of each element, one can derive the pattern of visible pixels as seen from the sweet spot of one view. As discussed in Sect. 3.1.2, this pattern behaves as a sub-sampling mask. As an example, the mask of a 24-view 3D display is shown in Fig. 18. The spectrum of this mask is shown in Fig. 28a. Each of the peaks in this spectrum corresponds to a source of aliasing. In order to avoid aliasing (Moiré artifacts), a filter has to be designed in such a way that its passband does not overlap with any of its copies generated by moving its center to any of those aliasing sources. It is possible that there are several different separable filters that can be used as antialiasing filters for this display, as shown in Fig. 28b. Each of those filters will perform proper antialiasing, but due to different shapes the visual quality of displayed images will be different. Which separable filter would yield best visual results depends on the content. The experiments presented in [70] suggest that for textual information such as subtitles, filters with a passband close to square perform better than filters with elongated passbands. For designing 1D filters with the desired cut-off frequencies, the windowing technique with the Kaiser window of length 24 can be used (see also Sect. 4.2.1). As an example, the magnitude responses (−6 dB contour) of several separable 2D filters suitable for the said 24-view display are shown in Fig. 28c.

4.2.3 Passband Approximation with a Tunable Filter

The results in [70] suggest that the filter that fully suppresses aliasing does not always give the best perceptual quality. Some people prefer sharper-looking images at the expense of some Moiré artifacts. In order to allow the user to control the

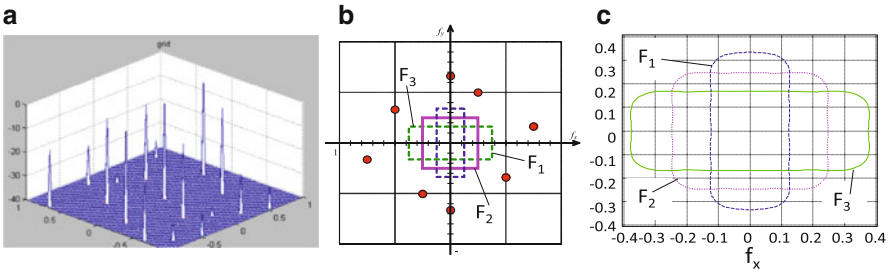


Fig. 28 2D separable filter: (**a**) spectrum of sub-sampling pattern for one view, (**b**) possible solutions for optimal antialiasing filters and (**c**) magnitude responses of these filters, −6 dB contour plots

antialiasing process according to his/her own preferences, one can design a set of tunable filters which depend on two parameters; apparent depth and desired sharpness. The sharpness parameter is expressed in terms of distortion-to-signal ratio, which is expected to affect the visibility of aliasing in perceptually linear fashion, regardless of the apparent depth.

An artifact mitigation framework which uses 1D tunable filters is proposed in [71]. It allows the user to specify the percentage of visible distortion over the original signal. The algorithm does the necessary processing to maintain the distortions within the selected limit, taking into account the display passband for different disparity values. It consists of three modules (Fig. 29) which are; offline processing module where the display is measured, display passband profile which is stored in a non-volatile memory and real-time processing module which filters the input image according to its apparent depth and selected distortion limits. During the measurements in the offline processing module, one derives the passband of the display for a range of disparity values as explained earlier. Each passband is approximated by a rectangle. The output from the module is stored in two tables. One table contains the height of the equivalent passband for various disparity values and levels of distortion, and the other table, the corresponding width of the passband. The real-time processing module uses these two tables to design the optimal filter for the input image. The system expects that the content (input to the framework) is stored in image-plus-depth format. The disparity value is used to select the corresponding column in each passband table. The user can set the value of the desired distortion level. This parameter is called “3D-sharpness” since it controls the tradeoff between visibilities of details versus visibility of Moiré artifacts. The value of “3D-sharpness” is used to select the corresponding row of each table. The values in the selected cells give the desired vertical and horizontal cut-off frequencies of an anti-aliasing filter. These cut-off frequencies are used for designing the filters.

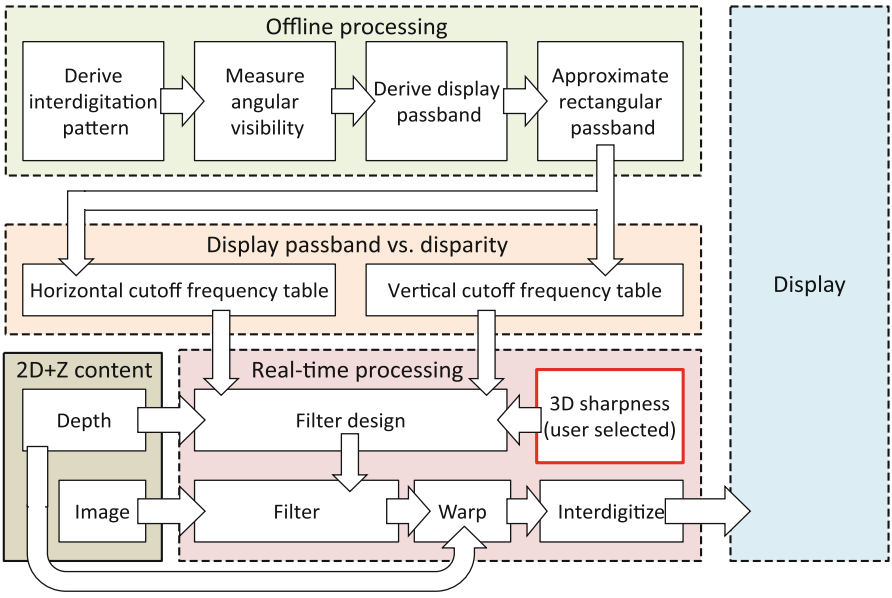


Fig. 29 Passband-optimization framework with tunable "3D sharpness" level

4.3 Content Optimization

4.3.1 Crosstalk Mitigation

In [67] Konradet et al. proposed a pre-compensation algorithm for reducing the crosstalk in stereoscopic displays. However, their approach is not suitable for multiview displays since pre-compensation mitigates the effect for a certain observation angle only, while amplifying it for other angles. As multiview displays are intended for many observers, it is desirable to mitigate the ghosting artifacts for all observation angles simultaneously. The straightforward approach to mitigate the crosstalk is to smooth the scene in horizontal direction, where the level of smoothing depends on the amount of the parallax (i.e. disparity) [80]. For a scene in image-plus-depth format this corresponds to smoothing of the 2D image, with the level of smoothing depending on the absolute depth values of the pixels.

Two algorithms for crosstalk mitigation for multiview displays, which can be implemented in a graphics processing unit (GPU), are discussed in [51]. The first algorithm employs pre-filtering of the 2D image before using it as a texture on the mesh. It uses eight filters for the whole range of depth values. Eight masks are prepared, passing a different range of depth values in accordance with the distance from the screen. Each mask is applied to the corresponding filtered image and the result is blended together in the accumulation buffer. The algorithm is implemented using the CUDA library [81].

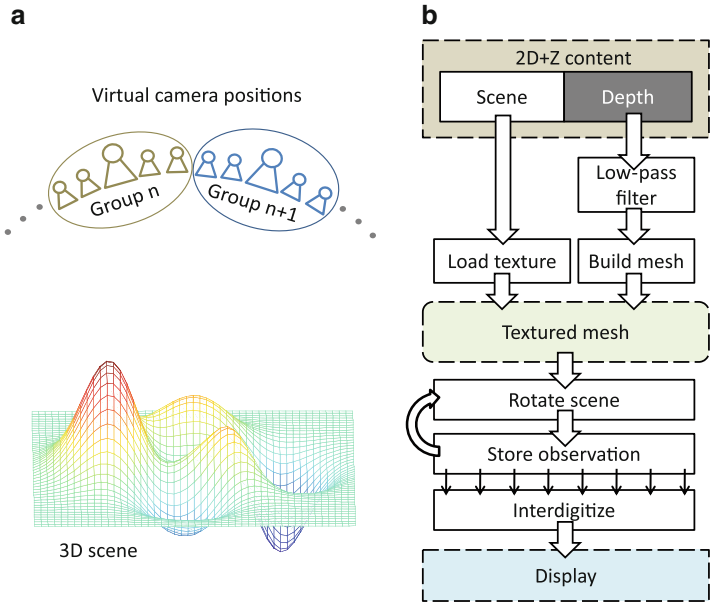


Fig. 30 Crosstalk mitigation with pre-filtering: (**a**) position of the extra observation points in respect to the original ones and (**b**) block diagram of the algorithm

The second algorithm uses an image scattering technique for crosstalk mitigation. It works by blending extra observations with the ones needed for the multiview display. Around each observation point used in previous approach, observation points are added at equal angles. The observation points are grouped as shown in Fig. 30a. The images rendered from a group of observation points are blended together in a single image which is mapped to the sub-pixels belonging to one view of the screen. The algorithm works in a similar way to the previous one, with the exception that instead of pre-filtering the texture; additional observations are rendered as illustrated in Fig. 30b.

4.3.2 Repurposing

Excessive disparity is a problem most often found in 3D content which is created for one display size and is observed in another. Adapting the size and disparity of 3D content to fit a given 3D display is known as *content repurposing*.

In [82] an algorithm for content repurposing on a mobile device is discussed. An important requirement for such an algorithm is that it can be used for real-time repurposing. Unfortunately, commonly used repurposing algorithms such as virtual view generation of non-liner disparity correction [46] are too computationally expensive to be used for real-time conversion on a contemporary portable device. In

3

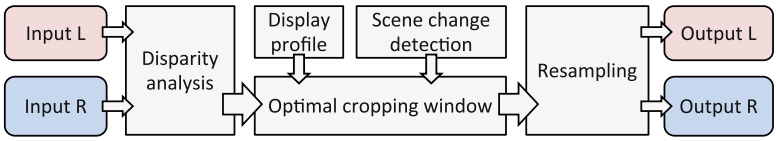


Fig. 31 Block diagram of an algorithm for fast 3D content repurposing

order to simplify the computation the algorithm uses *horizontal image translation* (HIT) which involves finding the size and position of a scaling window. In the HIT-based repurposing algorithm one first finds the disparity of the source video, then finds the optimal cropping and scaling parameters and then performs the actual image resampling. Having a dense disparity map is not critically important for performing HIT; it is enough to know the parameters of the disparity distribution, such as minimum, maximum and mean disparity.

The algorithm consists of five stages, as shown in Fig. 31. First, the comfort disparity range of the display is derived. Then the disparity range of the input content is calculated. Based on the estimated input and desired output disparity ranges, the algorithm derives the optimal scale of the cropping window which would yield the targeted disparity range and minimize the area of cropped and letterboxed content. Once the rescaling and cropping parameters are known we perform the resampling procedure with a desired, perceptually optimal performance in the frequency domain. More details on the algorithm performance can be found in [82].

5 Conclusions

Stereoscopic displays are meant to recreate a scene in three dimensions. Due to technical limitations of today's displays, some visual features of the scene are lost. The differences between the original scene and the reproduced one are interpreted by the HVS as artifacts. The missing information cannot be fully reconstructed, but due to absence of visual reference it is possible to make the distortions less visible. This raises the need of developing optimization techniques aimed at decreasing the visibility of artifacts on a 3D display. Such techniques require knowledge of both human vision and display design. One needs to know the important visual properties of a 3D scene and the relevant display properties that allow the scene to be shown in 3D.

This chapter presented signal processing techniques for optimized 3D scene visualization on contemporary 3D displays. Methods for deriving visual properties of 3D displays, predicting the visibility of artifacts, and visual optimization of 3D content were discussed. The presented methods for measuring the quality of a display are suitable for a large class of displays including glasses-enabled displays, portable autostereoscopic displays, and dual-view and multiview large

auto-stereoscopic displays. The proposed methods for visual optimization address issues such as observation position, head pose, view-multiplexing, and excessive disparity range and are effective for decreasing the visibility of the most common artifacts, experienced on stereoscopic displays, namely ghosting, Moiré and hyper-stereopsis.

References

1. B. A. Wandel, Foundations of Vision, Sunderland, Massachusetts, USA: Sinauer Associates, Inc, 1995.
2. I. P. Howard, and B. J. Rogers, Binocular Vision and Stereopsis, New York: Oxford University Press, 1995.
3. D. Chandler, "Visual Perception (Introductory Notes for Media Theory Students," MSC portal site, University of Wales, Aberystwyth, 2008. [Online]. Available: http://www.aber.ac.uk/media/sections/image.html.
4. S. Pastoor, "Human factors of 3D imaging: Results of recent research at Heinrich- Hertz-Institut Berlin," in *2nd International Display Workshop*, Hamamatsu, 1995.
5. M. Wexler and J. Boxtel, (2005) "Depth perception by the active observer," *Trends in Cognitive Sciences,* vol. 9, pp. 431–438.
6. B. Julesz, Foundations of Cyclopean Perception, Chicago: The University of Chicago Press, 1971.
7. E. Stoykova, A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, and X. Zabulis, "3-D Time-Varying Scene Capture Technologies—A Survey," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 17, no. 11, pp. 1568–1586, 2007.
8. P.-S. Tsai, J. Cryer, and M. Shah, "Shape-from-shading: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 21, no. 8, pp. 690–706 , 1999.
9. T. Lindeberg and J. Garding, "Shape from texture from a multi-scale," in *ICCV*, 1993.
10. M. Subbarao and G. Surya, "Depth from Defocus: A Spatial Domain Approach," *International Journal of Computer Vision,* vol. 13, pp. 271–294, 1994.
11. A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.
12. M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, no. 1, pp. 34–58, 2002.
13. H. Sidenbladh, M. Black, and L. Sigal, "Implicit probabilistic models of Human Motion for Synthesis and Tracking," in *European Conference on Computer Vision*, 2002.
14. S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," in *Proc. Comput. Vis. and Pattern Recognit. (CVPR2006)*, 2006.
15. B. L. Stann, A. Abou-Auf, S. Frankel, M. M. Giza, W. Potter, W. C. Ruff, P. H. Shen, D. R. Simon, M. R. Stead, Z. G. Sztankay, and L. F. Lester, "Research progess on scannerless ladar systems using a laser diode transmitter and FM/cw radar principles," in *Laser Radar Technology and Applications VI*, 2001.
16. U. Schnars and J. W., (1994) "Direct recording of holograms by a CCD target and numerical reconstructions," *Applied Optics,* vol. 33, no. 2, pp. 179–181.
17. A. Alatan, Y. Yemez, U. Gudukbay, X. Zabulis, K. Muller, C. Erdem, and A. Weigel, "Scene Representation Technologies for 3DTV—A Survey," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 17, no. 11, pp. 1587–1605, Nov. 2007.
18. M. Halle, "Multiple Viewpoint Rendering," in *Proceedings of the 25th annual Conference on Computer Graphics and Interactive Techniques*, 1998.

19. R. Hartly and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed., Cambridge University Press, 2006.
20. A. Smolic, K. Mueller, N. Stefanovski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, "Coding Algorithms for 3DTV - A Survey," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 17, no. 11, pp. 1606–1621, Nov. 2007.
21. C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, and I. Sexton, "An evolutionary and optimized approach on 3D-TV," in *Int. Broadcast Conf.*, Amsterdam, The Netherlands, 2002.
22. C. Fehn, N. Atzpadin, M. Muller, O. Schreer, A. Smolic, R. Tanger, and P. Kauff, "An Advanced 3DTV Concept Providing Interoperability and Scalability for a Wide Range of Multi-Baseline Geometries," in *2006 IEEE International Conference on Image Processing*, 2006.
23. R. Fernando and M. J. Kilgars, The Cg Tutorial, The Definitive Guide to Programmable Real-Time Graphics, Addison-Wesley, 2006.
24. J. Lee, "Hacking the Nintendo Wii Remote," *Pervasive Computing, IEEE,* vol. 7, no. 3, pp. 39–45, 2008.
25. K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Trans. Graph.,* vol. 23, no. 3, p. 804–813, 2004.
26. M. Saymta, S. Isikman, and H. Urey, "Scanning Led Array Based Volumetric Display," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2008.
27. S. Pastoor, "3D Displays," in *3D Video Communication*, O. Scheer, P. Kauff and T. Sikora, Eds., Chichester, West Sussex, Wiley, 2005, pp. 235–260.
28. P. Surman, K Hopf, I Sexton, W K Lee, R Bates "Solving the 3D problem - The history and development of viable domestic 3-dimensional video displays," in *Three-Dimensional Television: Capture, Transmission, and Display*, H. Ozaktas and L. Onural, Eds., Springer Verlag, 2007.
29. P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. von Kopylow, "A Survey of 3DTV Displays: Techniques and Technologies," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 17, no. 11, pp. 1647–1658, Nov. 2007.
30. H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the Art in Stereoscopic and Autostereoscopic Displays," *Proceedings of the IEEE,* vol. 99, no. 4, pp. 540–555., 2011.
31. L. Onural, T. Sikora, J. Ostermann, A. Smolic, M. R. Civanar, and J. Watson, "An Assessment of 3DTV Technologies," in *NAB Broadcast Engineering Conference Proceedings*, Las Vegas, USA, 2006.
32. H. Jorke, H. Simon, and M. Fritz, "Advanced stereo projection using interference filters," *J. Soc. Inf. Display, ,* vol. 17, no. 5, pp. 407–410, 2009.
33. Toshiba Europe GmbH, "55ZL2 - 3D without glasses," Toshiba, Jan 2012. [Online]. Available: http://eu.consumer.toshiba.eu/en/products/tv/55ZL2. [Accessed June 2012].
34. W. L. IJzerman, S. T. de Zwart, and T. Dekker, (2005) "Design of 2D/3D switchable displays," *Proc. of the SID,* vol. 36, no. 1, pp. 98–101.
35. C. van Berkel and J. Clarke, "Characterisation and optimisation of 3D-LCD module design," in *Stereoscopic Displays and Virtual Reality Systems IV*, San Jose, 1997.
36. W. Tzschoppe, T. Brueggert, M. Klipstein, I. Relke, and U. Hofmann, "Arrangement for two-or-three-dimensional display". US Patent 2006/0192908, 31 Aug. 2006.
37. M. Kristoffersen, M. J. Sykora, and J. Schultz, "Stretched filom for stereoscopic 3D display". US Patent 7,750,983, 6 July 2010.
38. N. Dodgson, "Autostereoscopic 3D Displays," *Computer,* vol. 38, no. 8, pp. 31–36, Aug. 2005.
39. A. Gotchev, B. G. Akar, T. Capin, D. Strohmeier, and A. Boev, "Three-Dimensional Media for Mobile Devices," *Proceedings of the IEEE,* vol. 99, no. 4, pp. 708–737, 2011.
40. J. Konrad and P. Angiel, "Subsampling models and anti-alias filters for 3-D automultiscopic displays," *IEEE Trans. Image Processing,* vol. 15, no. 1, pp. 128–140, 2006.
41. C. van Berkel, "Lenticular screen adaptor". US Patent 6801243, 5 Oct. 2004.
42. V. Saveljev, J.-Y. Son, B. Javidi, S.-K. Kim, and D.-S. Kim, "Moiré minimization condition in three-dimensional image displays," *Display Technology,* vol. 1, pp. 347–353, 2005.

43. C. N. Moller and A. R. L. Travis, "Correcting interperspective aliasing in autostereoscopic displays," *, IEEE Trans. Visual Comput. Graphics,* vol. 11, no. 2, pp. 228–236, 2005.
44. D. Hoffman, A. Girshick, K. Akeley, and M. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision,* vol. 8, no. 3, pp. 1–30, 2008.
45. S. K. Nayar, V. Branzoi, and T. E. Boult, "Programmable Imaging Using a Digital Micromirror Array," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2004.
46. M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear Disparity Mapping for Stereoscopic 3D," *ACM Transactions on Graphics (Proc. SIGGRAPH),* vol. (in press), 2010.
47. W. IJsselsteijn, P. Seuntiens, and L. Meesters, (2005) "Human factors of 3D displays," In *3D Video Communication*, Scheer, Kauff and Sikora, Eds., Wiley, pp. 219–233.
48. M. Halle, "Autostereoscopic displays and computer graphics," in *International Conference on Computer Graphics and Interactive Techniques*, 2005.
49. A. Boev, R. Bregovic, and A. Gotchev, "Visual-quality evaluation methodology for multiview displays," *Displays,* vol. 33, no. 2, pp. 103–112, April 2012.
50. A. Boev, D. Hollosi, A. Gotchev, and K. Egiazarian, "Classification and simulation of stereoscopic artifacts in mobile 3DTV content," in *Stereoscopic Displays and Applications XX, Proc. SPIE 7237*, 2009.
51. A. Boev, K. Raunio, A. Gotchev, and K. Egiazarian, "GPU-based algorithms for optimized visualization and crosstalk mitigation on a multiview display," in *Proc. SPIE 6803*, 2008.
52. F. Kooi and A. Toet, "Visual comfort of binocular and 3D displays," *Displays,* vol. 25, no. 2–3, pp. 99–108, 2004.
53. A. Jain and J. Konrad, "Crosstalk on automultiscopic 3-D displays: Blessing in disguise?," in *Stereoscopic Displays and Applications XVIII, IS&T/SPIE Electronig Imaging*, San Jose, CA, 2007.
54. M. Salmimaa and T. Järvenpää, "Optical characterization of autostereoscopic 3-D displays," J. Soc. Inf. Display, vol. 16, no. 825, 2008.
55. J. Häkkinen, J. Takatalo, M. Kilpeläinen, M. Salmimaa and G. Nyman, "Determining limits to avoid double vision in an autostereoscopic display: Disparity and image element width," J. Soc. Inf. Display, vol. 17, no. 433, 2009.
56. E. D. Montag and M. D. Fairchild, "Fundamentals of Human Vision and Vision Modelling," in *Digital Video Image Quality and Perceptual Coding*, H. R. Wu and K. H. Rao, Eds., Boca Raton, FL, CRC Press, 2006, pp. 45–81.
57. A. Boev and A. Gotchev, "Comparative study of autostereoscopic displays for mobile devices," in *Multimedia on Mobile Devices 2011, Proc. SPIE 7881*, 2011.
58. B. T. Bakus, M. S. Banks, R. van Ee, and J. A. Crowell, "Horizontal and vertical disparity, eye position, and stereoscopic slant perception," *Vision Research,* vol. 39, pp. 1143–1170, 1999.
59. P. Boher, T. Leroux, T. Bignon and V. Collomb-Patton, "A new way to characterize autostereoscopic 3D displays using Fourier optics instrument," in *Stereoscopic Displays and Applications XX, SPIE 7237, 72370Z*, 2009.
60. M. Salmimaa and T. Järvenpää, "3-D crosstalk and luminance uniformity from angular luminance profiles of multiview autostereoscopic 3-D displays," *Soc. Inf. Display,* vol. 16, p. 1033, 2008.
61. A. Boev, K. Raunio, M. Georgiev, A. Gotchev and K. Egiazarian, "OpenGL-based Control of Semi-Active 3D Display," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, Istanbul, Turkey, 2008.
62. S. Uehara, T. Hiroya, H. Kusanagi, K. Shigemura and H. Asada, "1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement," in *Stereoscopic displays and applications XIX*, 2008.
63. A. Boev, R. Bregovic and A. Gotchev, (2010) "Measuring and modeling per-element angular visibilty in multiview displays," *Special issue on 3D displays, Journal of Society for Information Display,* vol. 18, no. 9, pp. 686–697.

64. P. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," in *ACM Siggraph*, 1997.
65. A. Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4D-vision," in *SPIE Photonics West 2002: Electronic Imaging*, 2002.
66. S. Winkler, Digital Video Quality, John Wiley & Sons, 2005.
67. J. Konrad, B. Lacotte, and E. Dubois, "Cancellation of image crosstalk in time-sequential displays of stereoscopic video," *IEEE Trans. Image Process.,* vol. 9, pp. 897–908, May 2000.
68. M. Zwicker, W. Matusik, F. Durand, H. Pfister, and C. Forlines, "Antialiasing for automultiscopic 3D displays," in *ACM SIGGRAPH 2006*, Boston, Massachusetts, 2006.
69. A. Boev, R. Bregovic, and A. Gotchev, "Methodology for design of anti-aliasing filters for autostereoscopic displays," *Special issue on Advanced Techniques on Multirate Signal Processing for Digital Information Processing, Journal of IET Signal Processing,* vol. 5, no. 3, pp. 333–343, June 2010.
70. A. Boev, R. Bregovic, A. Gotchev, and K. Egiazarian, "Anti-aliasing filtering of 2D images for multi-view auto-stereoscopic displays," in *The 2009 International Workshop on Local and Non-Local Approximation in Image Processing, LNLA 2009*, Helsinki, Finland, 2009.
71. A. Boev, R. Bregovic, and A. Gotchev, "Design of tuneable anti-aliasing filters for multiview displays," in *Stereoscopic Displays and Applications XXII, Proc. SPIE 7863*, 2011.
72. R. Brar, P. Surman, I. Sexton, R. Bates, W. Lee, K. Hopf, F. Neumann, S. Day, and E. Willman, "Laser-Based Head-Tracked 3D Display Research," *Display Technology, Journal of,* vol. 6, no. 10, pp. 531–543, 2010.
73. K. Hopf, F. Neumann, and D. Przewozny, "Multi-user eye tracking suitable for 3D display applications," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, 2011.
74. A. Boev, M. Goergiev, A. Gotchev, N. Daskalov, and K. Egiazarian, "Optimized visualization of stereo images on an OMAP platform with integrated parallax barrier auto-stereoscopic display," in *17th European Signal Conference, EUSIPCO 2009*, Glasgow, Scotland, 2009.
75. V. Uzunov, A. Gotchev, K. Egiazarian, and J. Astola, "Face Detection by Optimal Atomic Decomposition," in *SPIE Optics and Photonics 2005: Algorithms, Architectures, and Devices and Mathematical Methods, Mathematical Methods in Pattern and Image Analysis*, San Diego, California, USA, 2005.
76. N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis,* vol. 10, no. 3, pp. 234–253, May 2001.
77. H. Essaky Sankaran, A. Gotchev, K. Egiazarian, and J. Astola, "Complex wavelets versus Gabor wavelets for facial feature extraction: a comparative study," in *Proc. SPIE Image processing : algorithms and systems IV, Vol. 5672*, San Jose, CA, 2005.
78. A. Boev, M. Georgiev, A. Gotchev, and K. Egiazarian, "Optimized single-viewer mode of multiview autostereoscopic display," in *Proc. of 16th European Signal Conference EUSIPCO 2008*, Lausanne, Switzerland, 2008.
79. S. K. Mitra, Digital signal processing: A computer based approach, 3 ed., New York: McGraw-Hill, 2005.
80. M. W. Halle, "Holographic stereograms as discrete imaging systems," in *Practical Holography VIII*, San Jose, CA.
81. V. Podlozhnyuk, "Image Convolution with CUDA, white paper," Nvidia Corp, June 2007. [Online]. Available: http://developer.download.nvidia.com. [Accessed June 2012].
82. A. Karaoglu, B. H. Lee, W.-S. Cheong, A. Boev, and A. Gotchev, "Fast repurposing of high-resolution stereo video content for mobile use," in *Real-Time Image and Video Processing 2012,*, Brussels, Belgium, 2012.