Immersive GPU-Driven Biological Adaptive Stereoscopic Rendering

Sammy Rogmans*,†, Maarten Dumont*, Gauthier Lafruit† and Philippe Bekaert*
*Hasselt University – tUL – IBBT, Expertise centre for Digital Media
Wetenschapspark 2,
BE-3590 Diepenbeek, Belgium
e-mail: {firstname.name}@uhasselt.be
†Multimedia Group, IMEC
Kapeldreef 75,
BE-3001 Leuven, Belgium
e-mail: {firstname.name}@imec.be

***Abstract*—In this paper, we want to sensitize 3D content developers and researchers of broadening their scope of parameters they take into account for generating 3D content. State-of-the-art perceptual research has already shown that monocular visual cues highly contribute to the very fundamentals of 3D perception, and binocular ones are merely linked to them in order to create a rich depth experience. In this context, we present an overview of the research concerning our teleconferencing system that is able to recreate biological stereoscopic input, without loosing consistency in all related monocular cues such as accommodation, occlusion, size (gradient), motion parallax, texture gradient and linear perspective. The system adapts in real-time by doing both GPU-driven analysis and rendering, based on the physical parameters of the system user.**

I. Introduction

Lately there is a complete revival of 3D hardware, often using over 100 Hz LED televisions or monitors with integrated synchronization transmitters for active shutterglasses, which are available as commodity off-the-shelf products, while movie theaters are also rapidly making the transition to 3D using digital projectors. It is a sort of transitional technology leading to autostereoscopic displays, which even present more than two synchronized images, to set up 3D perception without using any special glasses whatsoever.

The 3D revival is making the production of 3D content interesting again, nonetheless researchers and developers are motivated not to make the same mistakes as in former attempts to commercialize 3D perception. A lot of attention is given to avoid any visual fatigue and aberrations in the way people perceive the stereoscopic images. This attention is mainly focussed on vergence and stereopsis, being the binocular biological processes in the brain that directly lead to rich 3D perception. However, recently leading vision-perceptual researchers have found more and more important factors which all contribute to a more fundamental 3D experience that has a comfortable and natural feel to it, even being able to create significant 3D impressions without using stereoscopy at all [1]. Our experiments further acknowledge these findings, and we therefore present an teleconfercing system (see Fig. 1) that is able to adaptively render stereoscopic images with the goal of being maximally consistent with the biological processes of 3D perception in the brain, resulting in a true immersive, comfortable and natural depth experience.

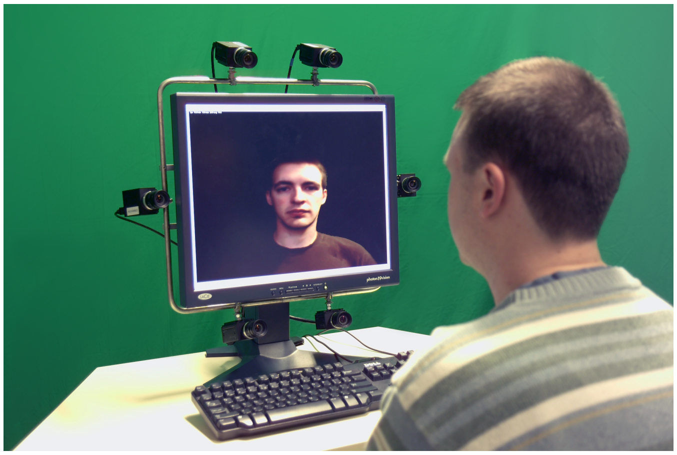

Fig. 1. Example setup of our immersive one-to-one teleconferecing system, using multiple cameras to allow virtual camera image generation.

II. Visual Depth Cues

It is a common misperception that perceiving natural 3D is only the consequence of synchronized visual input to both eyes. Perceiving depth in a natural and comfortable way is a highly complex biological process that occurs within the brain, which involves fusing and interpreting different visual cues of the human vision system. As depicted in Fig. 2, visual cues can be subdivided in two distinct groups – i.e. the monocular and binocular cues – which relate to providing additional depth information from one-eye individual and two-eye simultaneous visual input respectively. Lately more and more researchers are becoming aware that consistent monocular input is at least equally – if not more – important than correct binocular input in synthetically trying to recreate natural 3D perception, e.g. using active shutter glasses or autostereoscopic displays.

Binocular cues should indeed receive proper attention, nonetheless the amount of monocular cues are far greater,

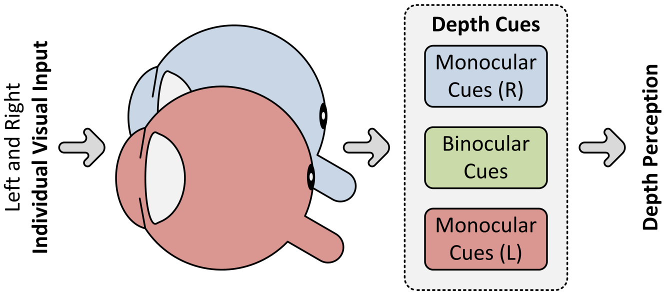

Fig. 2. Schematical representation concerning the different depth cues of the human vision system that lead to natural depth perception.

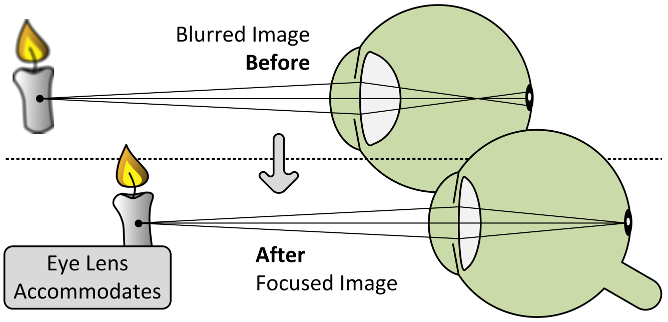

Fig. 3. Accommodation of the eye lens after a change in visual fixation, placing the focal point upon the retina and macula.

which implicates the higher probability of an inconsistent visual cue of this type. Binocular cues contribute to the richness of depth experience, however they are inevitably linked to the monocular ones by the biological processes in the brain. As a result, a single inconsistency in one of the monocular cues may lead to the annihilation of realistic and comfortable synthetic 3D perception.

A. Monocular Cues

There are nine distinct types of monocular cues in total. To easily understand the monocular cue types, a minimal knowledge of the eye anatomy is required, basically dissecting the front of the eye in the cornea, iris and pupil, the lens within the eye, and the retina, macula and fovea at the back, connecting to the optical nerve. Monocular cues are based on individual visual input from a single eye, and are therefore still active when one eye should be shut or disabled. The different types of monocular cues are:

1) **Accommodation**: When changing visual fixation from one to another object, the image appears briefly blurred for the brain, since the focal point of the light rays entering the pupil do not coincide with the retina and fovea, i.e. the central part of the macula which contributes most significantly to clear vision. The brain quickly responds to this, and the eye lens accommodates accordingly by intraocular muscles to focus the image (see Fig. 3). The movement and position of the related eye muscles provide with oculomotor feedback to the brain, containing both relative and absolute depth information.
2) **Occlusion**: Also called interposition or overlapping, occlusion is the most trivial and apparent visual depth cue that provides with relative depth information. Objects that seem to occlude other ones, are interpreted as being closer. It is a simple but powerful cue that is able to make or break the entire natural 3D perception, e.g. subtitles in movies are often rendered on the screen plane while still occluding scene content that pops out. Many people expierence this phenomenom as disturbing and some even get sick of this visual cue being inconsistent.
3) **Size and Size Gradient**: As the brain aquires prior knowledge about the size of certain objects (e.g. cars, houses, etc.), estimating depth in an absolute manner becomes possible and more reliable. Furthermore, if one of multiple similar objects appears smaller, the brain will hence interprete that object as being relatively farther away. Analogous to this, objects shrinking in size appear to be moving farther away, while objects that increase in size will be perceived as coming closer.
4) **Motion Parallax**: Even a single eye can convincingly perceive depth if the head is moved perpendicular to the viewing direction. Objects that are closer will exhibit more parallax, i.e. the amount of visual shift between both positions is larger, while objects that are far away will almost appear to not move at all.
5) **Texture Gradient**: The more detail that can be seen, the closer an object will be interpreted. This cue should not be mistaken with accommodation, as it provides pictoral instead of oculomotor feedback to the brain. Nevertheless, they are closely linked to each other since blur is the natural drive for accommodation. Manipulated image blur can therefore drastically affect the perceived distance and even size of objects. The state-of-the-art research in [1] describes this in more extensive detail.
6) **Shades and Shadows**: The brain generally tends to assume light always comes from above, due to the fact that most of the light is provided by the sun. Although whenever the position of an possibly artificial light source is quite clear, the shades need to be consistent with the given light position. The shades and casted shadows of an object therefore provide with extra relief and depth information.
7) **Linear Perspective**: This visual cue enables the capability of recognizing planes and estimating vanishing points, e.g. parallel lines of a road that eventually meet in the horizon, as a consequence of the gradient size visual cue. Because of its close relation, some people have argued that it is not a separate cue [2]. It does not necessarily provide with absolute depth information, but provides insight in the relative structure of a scene.
8) **Relative Height**: As the name already gives away, this cue provides with relative depth information since objects that are smaller and closer to the horizon are observed as being farther away.

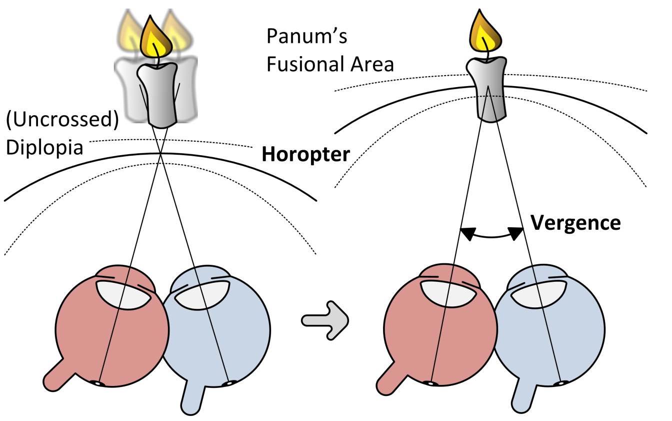

Fig. 4. The process of vergence symetrically converges the two eyes, resulting in the capablity of fusing the binocular input.

9) **Aerial Perspective**: Due to water and dust particles in the atmosphere, objects that are more in the background and open air, will appear more hazy since more particles are in between the object and the viewer. It is not to be confused with texture gradient, as aerial perspective also decreases the contrast and color saturation.

B. Binocular Cues

In contrast with the various monocular cues, only two types of binocular visual depth cues can be distinguished. They require consistent visual input from both eyes simultaneously, and are therefore the most fragile and complex, but are able to provide a truly rich 3D experience whenever they are enabled. The different binocular cue types are:

1) **Vergence**: Visual fixation on an object with both eyes requires them to appropriately converge and rotate by the use of extraocular muscles. This process is guided by the brain, and is stimulated by four factors, i.e. accommodative (individual eye focus), tonic (concious use of neck muscles), proximal (awareness of proximity) and fusional (desire for single vision) convergence. Analogous to the accommodation, this cue also provides with oculomotor feedback for absolute depth. However, as accommodation induces the desire to converge, there is a significant link between them. Since the sixties, research has already shown that the process of vergence results over two thirds from accommodative convergence [3], and for only one third from tonic, proximal and fusional convergence, making the link between accommodation and vergence very powerful.
2) **Stereopsis**: As shown in Fig. 4, the two eyes converge symmetrically, causing light rays from points in space to be captured by corresponding photoreceptive areas in the two retinas. Such a point is said to have zero (angular) disparity. Moreover, the locus of these points is called the horopter. Points or objects in front or beyond the horopter will cause crossed or uncrossed disparities, which the brain is able to interprete as rich continious depth information. Furthermore, the brain is only capable of fusing binocular input within the vicinity of the horopter, which is often reffered to as Panum's fusional area (see Fig. 4). Objects outside this area cause crossed or uncrossed diplopia – i.e. double vision – because, identical to most monocular cues, stereopsis only provides with relative depth information, i.e. within Panum's fusional area.

Since accommodation, vergence and familiar size are the only depth cues that provide with absolute depth information, the importance of monocular relative visual depth cues is often drastically underrated in synthetically sustaining natural and rich depth perception.The probability of them destroying the genuine feel of the 3D perception is inevitably very high because of the closely linked biological processes, and should therefore receive the proper attention when rendering stereoscopic images. In extremis, a stereoscopic 3D rendering should be fully adapted to an individual viewer, dependig on his or her physical characteristics and viewing location.

III. Biological Adaptive Stereoscopy

We developed a system that consists out of a number of parallel and pipelined processing modules – which take a huge amount of computations – that continiously analyzes the participant's parameters and dynamically adjusts the stereoscopic rendering, based on the biological processes of 3D perception. It sustains consistency in all relevant visual depth cues by also exploiting the power of the monocular cues. This provides a natural and comfortable way of perceiving depth, which maximizes the immersion of the participant.

To enable the real-time processing of both the multicamera analysis and biological adaptive stereoscopic rendering, we rely on the massive parallel computation power within a GPU, using traditional shaders and the CUDA language, according to specific mapping methodologies that optimize the utilization of the hardware.

A. Ultimate Immersion

In previous research, we developed a one-to-one teleconference system [4], [5] that uses a six-fold camera multiview setup to interpolate a virtual camera, as if it was located behind the screen and directly looking into the eyes of the participant, succesfully restoring eye contact. Our one-to-one teleconference system therefore provides the ideal context for experimenting with adaptive stereoscopy and the ultimate immersion, as there is only one participant for each monitor and the stereoscopic feed can be rendered on an individual level by using our existing virtual camera interpolation techniques.

While a detailed overview of the parallel and pipelined processes can be found in [6], the main processing module uses the distance from the user to the screen to adjust the virtual cameras in such a way that the other participant's eyes always converges to the screen, i.e. being equal to the accommodation distance. An apparant disadvantage of this technique is that the 3D experience of the rendered participant – and all scene content for that matter – is flattened around the screen plane, nevertheless stereopsis and vergence are still

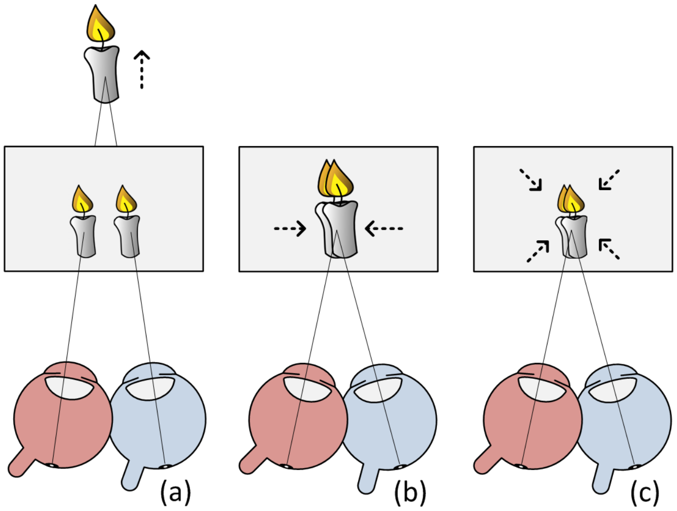

Fig. 5. Our system process where (a) vergence, that does not coincide with the accommodation, is corrected by (b) moving the virtual cameras to the given accommodation distance of the sytem user, and (c) exploiting the size gradient cue to give the appearance of depth motion.

enabled. As depicted in Fig. 5, we therefore exploit the size gradient, texture gradient and other monocular cues, as further explained in [7], to trick the brain in thinking the participant is actually moving closer or farther away, without ever destroying the important link between vergence and accommodation. This not only keeps the vergence cue in Panum's fusional area, but even within Percival's zone of viewing comfort, which is even a smaller area around the horopter that garantuees optimal 3D depth perception without any visual fatigue.

B. GPU-Driven Rendering

Continiously analyzing and adjusting the virtual camera images takes a tremendous amount of computations and is only feasible by using the vast amount of processing power available within GPUs. Our implementation is inherently parallel and therefore a good match for this parallel architecture, while optimizing between traditional shaders and the more novel CUDA general purpose language [8].

In [9], we have developed a smart analysis control loop to increase both speed and quality of the virtual camera interpolation, by determing and using the position of the participant towards the screen. The output of this intelligent process is also directly pipelined to the processing modules that are responsible for maintaining the vergence-accommodation link. Furthermore, we have developed algorithmic-specific mapping methodologies in [10], [11] and [12] to fully optimize both spacial and temporal utilization of the hardware, hence maximizing the execution efficiency.

C. Experimental Results

Considering normal-speed user movements, our control loop is able to succesfully stabilize the depth perception – by dynamically adjusting the vergence – right in the center of Percival's zone of comfort, resulting in very close to real world 3D perception. The monocular cues are furthermore accordingly exploited to trick the brain in seeing seemless movement in and out the screen. While doing this, the system is still able to run in real-time at 41 fps for a resolution of 800×600 pixels on an NVIDIA GeForce 8800GTX.

IV. Conclusion and Future Work

We have presented an overview of the research we performed concering our one-to-one teleconference system. By using a GPU-driven control loop that detects the position of the user towards the screen, we are able to render a biologically adapted stereo image in real-time, to maintain a maximum consistency in all relevant visual depth cues that contribute to a natural and comfortable 3D perception. We have noticed significant improvements in the immersion-level of the system and its sustainability. We therefore hope to have sensitized 3D content developers and researchers to take all visual cues, definitely the monocular ones, into account when generating future 3D content.

In the future, we are planning on performing valid user tests in collaboration with the human-computer interfacing research group. Furthermore, we will be making the transition to autostereoscopic displays, as we already have important results in rendering multiple virtual images.

References

[1] Robert T. Held, Emily A. Cooper, James F. O'Brien, and Martin S. Banks, "Using blur to affect perceived distance and size," *ACM Transactions on Graphics*, vol. 29, no. 2, pp. 19:1–16, March 2010.

[2] James E. Cutting, "How the eye measures reality and virtual reality," *Behavior Research Methods, Instruments & Computers*, vol. 29, no. 1, pp. 27–36, 1997.

[3] A. Hughes, "AC/A ratio," *British Journal of Ophthalmology*, vol. 51, no. 11, pp. 786–787, November 1967.

[4] Maarten Dumont, Steven Maesen, Sammy Rogmans, and Philippe Bekaert, "A prototype for practical eye-gaze corrected video chat on graphics hardware," in *SIGMAP*, Porto, Portugal, July 2008, pp. 236–243.

[5] Maarten Dumont, Sammy Rogmans, Steven Maesen, and Philippe Bekaert, "Optimized two-party video chat with restored eye-contact using graphics hardware," *Communications in Computer and Information Science*, vol. 48, pp. 358–372, November 2009.

[6] Maarten Dumont, Sammy Rogmans, Gauthier Lafruit, and Philippe Bekaert, "Immersive teleconferencing with natural 3d stereoscopic eye contact using gpu computing," in *3D Stereo Media*, Lige, Belgium, December 2009.

[7] Sammy Rogmans, Maarten Dumont, Gauthier Lafruit, and Philippe Bekaert, "Biological-aware stereoscopic rendering in free viewpoint technology using gpu computing," in *3DTV-CON*, Tampere, Finland, June 2010.

[8] Sammy Rogmans, Maarten Dumont, Gauthier Lafruit, and Philippe Bekaert, "Migrating real-time image-based rendering from traditional to next-gen GPGPU," in *3DTV-CON*, Potsdam, Germany, May 2009.

[9] Sammy Rogmans, Maarten Dumont, Tom Cuypers, Gauthier Lafruit, and Philippe Bekaert, "Complexity reduction of real-time depth scanning on graphics hardware," in *VISAPP*, Lisbon, Portugal, February 2009, pp. 547–550.

[10] Patrik Goorts, Sammy Rogmans, and Philippe Bekaert, "Optimal data distribution for versatile finite impulse response filtering on next-generation graphics hardware using CUDA," in *ICPADS*, Shenzhen, China, December 2009, pp. 300–307.

[11] Sammy Rogmans, Philippe Bekaert, and Gauthier Lafruit, "A high-level kernel transformation rule set for efficient caching on graphics hardware," in *SIGMAP*, Milan, Italy, July 2009.

[12] Patrik Goorts, Sammy Rogmans, and Philippe Bekaert, "Practical examples of gpu computing optimization principles," in *SIGMAP*, Athens, Greece, July 2010.