

# Classification and simulation of stereoscopic artifacts in mobile 3DTV content

Atanas Boev, Danilo Hollosi, Atanas Gotchev, Karen Egiazarian  
Institute of Signal Processing, Tampere University of Technology, P.O.Box 553, 33101 Tampere,  
Finland

## ABSTRACT

We identify, categorize and simulate artifacts which might occur during delivery of mobile stereoscopic video. We consider the stages of 3D video delivery dataflow: content creation, conversion to the desired format (multiview or dense-depth 3D video), coding/decoding, transmission, and visualization on 3D display. The 3D vision of humans works by assessing various depth cues – accommodation, binocular depth cues, pictorial cues and motion parallax. As a consequence any artifact which modifies these cues will impair the quality of a 3D scene.

The perceptibility of each artifact can be estimated through subjective tests. The material for such tests should exhibit various artifacts with different amounts of impairment. We present a system for simulation of such artifacts. The artifacts are organized in groups with similar origins, and each group is simulated by a block in a simulation channel. The channel consists of the following blocks: simulation of sensor limitations, simulation of geometric distortions as the ones caused by the camera optics, spatial and temporal misalignments between video channels, spatial and temporal artifacts caused by coding, transmission losses, and visualization artifacts. For the case of dense depth video representation, format conversion artifacts are added.

**Keywords:** mobile 3DTV, stereoscopic artifacts, stereoscopic quality, mobile 3D video, portable 3D displays

## 1. INTRODUCTION

Recently, most of the building blocks of an end-to-end mobile 3DTV system have reached maturity status. An ISO/MPEG multiview encoding standard developed as an amendment to H.264 AVC is being standardized<sup>1, 2</sup>. Various algorithms have been developed for the efficient transmission of video streams over wireless networks<sup>1, 3</sup>. There are 3D displays optimized for a mobile use<sup>4, 5, 6</sup>. While the core technologies have been developing, there is still much to be done to optimize the system to deliver the best possible visual output<sup>7, 8</sup>. Having a perceptually acceptable and high-quality 3D scene on a small display is a challenging task.

Estimation of the quality is the key factor in design and optimization of any visual content. All quality metrics aim at close approximation of the quality as perceived by the user. An ideal quality metric should have the following properties: a) perceptual – being related to the way human visual system (HVS) operates, b) objective – providing a numerical representation of the quality as perceived by the user, and c) reliable – being able to predict the perceptual quality for wide variety of content, as perceived by a large amount of users. Such metric is especially needed for stereoscopic 3D video, because stereoscopic artifacts would produce not only visually unpleasant results, but are also known to cause eye-strain general discomfort<sup>9</sup>. The previous works on quality of stereo images<sup>10, 11</sup> do not attempt to quantify the typical distortions that could occur in stereoscopic video sequence.

The first step towards objective quality estimation metric is to identify the artifacts, which could occur in which could arise in various usage scenarios involving stereoscopic content. Then, subjective tests should be performed, in which human observers would grade the perceptual quality of a variety of content. In this work, we attempt to identify the artifacts, which could occur in a mobile 3DTV system. We present a system which could introduce a set of stereoscopic artifacts to a given 3D video, thus ensuring repeatability of subjective quality experiments. In the next chapter, we discuss the “layered” nature of the human 3D vision. In chapter 3 we introduce a concept for broadcasting stereo-video over DVB-H channel, and describe which stages of such system can introduce artifacts. In chapter 4, we compare the delivery stages to the “layers” of 3D vision to build a classification of stereoscopic artifacts. In chapter 5 we present a framework for simulation of mobile 3DTV artifacts. Finally, chapter 7 describes the mobile 3DTV artifacts simulated by our framework.

## 2. PERCEPTION OF DEPTH

The human visual system is a set of separate subsystems, which operate together in a single process. It is known that spatial, color and motion information is transmitted to the brain using largely independent neural paths<sup>12</sup>. Vision in 3D, in turn, also consists of different “layers” which provide separate information about depth of the observer scene<sup>12, 13</sup>. This is true both for perception and cognition – on perceptual level there are separate visual mechanisms and neural paths, and on cognitive level there are separate families of depth cues, with varying importance from observer to observer<sup>12, 14</sup>. The depth cues used in different layers in human vision are shown in Figure 1 and are as follows:

- Accommodation – This is the ability of the eye to change the optical power of its lens in order to focus on objects at various distances. Accommodation is the primary depth cue for very short distances, where an object is hardly visible with two eyes. With the distance, the importance of this depth cue quickly decreases. However, the information from other depth-assessing systems is unconsciously used to correct the refraction power, to ensure clear image of the object being tracked. As a result, a discrepancy between accommodation and binocular depth cues leads to so called *accommodation-convergence rivalry*, which is a major limiting factor for stereoscopic displays.
- Binocular depth cues – These are a consequence of both eyes observing the scene at slightly different angles. The mechanism of binocular depth estimation has two parts – *vergence* and *stereopsis*. Vergence is the process, in which both eyes take a position which minimizes the difference of the visual information projected in both retinæ. The angle between the eyes is used as a depth cue. With the eyes converged on a point, stereopsis is the process which uses the residual disparity of the surrounding area for depth estimation relative to the point of convergence. Binocular depth cues are the ones most often associated with “3D cinema”. However, binocular vision is quite vulnerable to artifacts – lots of factors can lead to an “unnatural” stereo-pair being presented to the eyes. As HVS is not prepared to handle such information, binocular artifacts can lead to nausea and *simulator sickness* or *cyber sickness*<sup>9</sup>. It is worth saying, that around 5% of all people are “stereoscopically latent” and have difficulties assessing binocular depth cues<sup>11, 13</sup>. Such people have a perfect depth perception, only they rely mostly on depth cues coming from other visual “layers”.
- Pictorial cues – for longer distances, binocular depth cues become less important, and HVS relies on pictorial cues for depth assessment. These are depth cues that can be perceived even with a single eye – shadows, perspective lines, texture scaling. But even for medium distances, stereoscopically good image can be “ruined” if missing subtle pictorial details, and the scene exhibits *puppet theatre* or *cardboard effect* artifacts
- Motion parallax – this is the process in which the changing parallax of a moving object is used for estimating its depth and 3D shape. The same mechanism is used by insects, and is commonly known as “insect navigation”<sup>15</sup>. Artifacts in the temporal domain (e.g. motion blur, display persistence) will affect the motion parallax depth cues.

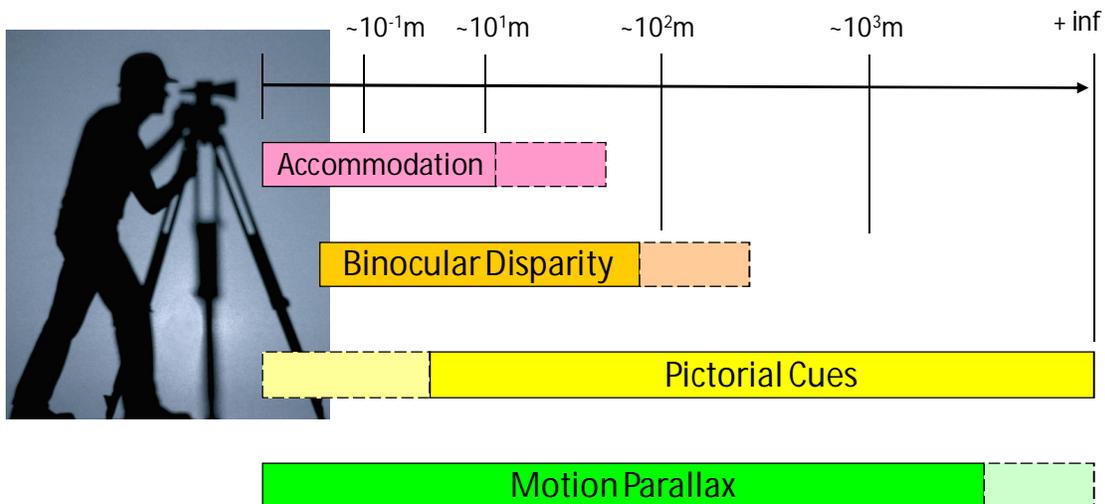


Figure 1, Depth perception as a set of separate visual “layers”

Experiments with so-called “random dot stereograms” show that binocular and monocular depth cues are independently perceived<sup>16</sup>. Furthermore, the first binocular cells (cells that react to a stimulus presented to either of the eyes) appear at a late stage of the visual pathways – the V1 area of brain cortex. At this stage, only the information extracted separately for each eye, is available to the brain for deduction of image disparity<sup>12</sup>. This observation has led to our assumption that “2D” (monoscopic) and “3D” (stereoscopic) artifacts would be independently perceived<sup>17</sup>. The planar “2D” artifacts, such as noise, ringing, etc. are thoroughly studied in the literature<sup>18</sup>. Here, we focus on artifacts which affect stereoscopic perception. However, due to the “layered” structure of HVS, binocular artifacts might be inherited from other visual “layers” – for example, *blockiness* is a “purely” monoscopic artifact, which still can destroy or modify an important binocular depth cue.

### 3. ARTIFACTS IN MOBILE 3DTV SYSTEM

The dictionary describes *artifact* as “something characteristic of or resulting from a human institution or activity”<sup>19</sup>. Non-natural processes, as is the case of transmitting a 3D scene representation over a communication channel, are source of artifacts.

One case of such transmission is a mobile 3DTV system, where a 3D (usually stereoscopic) video stream is broadcasted over the air and received on a portable device. We consider the following scenario of mobile 3DTV content delivery: stereoscopic video content is captured, encoded, encapsulated and then broadcast over mobile TV (DVB-H system) to be received, decoded and played by a DVB-H enabled portable device with autostereoscopic display. The data flow from creation to observation is shown in Figure 2.

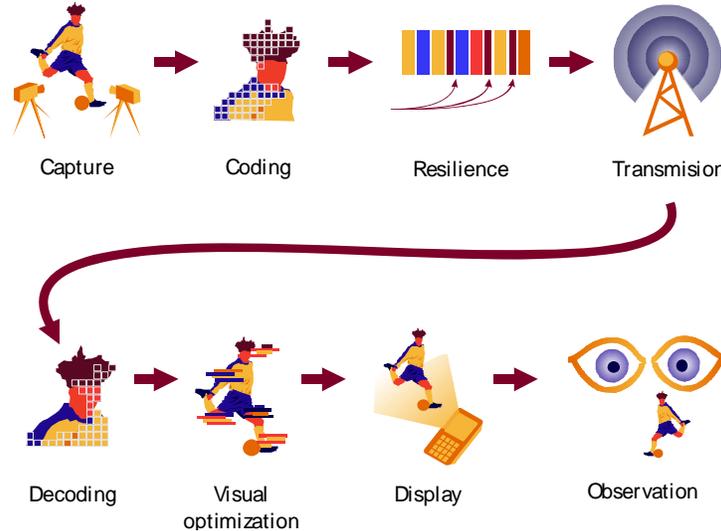


Figure 2, Data flow of mobile 3DTV content

The stages of the dataflow can create various artifacts as follows:

- Creation/capture – this is the stage, where there are three common approaches for capturing of 3D video. First, such video can be captured by two or more synchronized cameras in a multi-camera setting. Second, such content can be created from 2D video applying video processing methods. Third, video output can be augmented by depth information captured by another sensor. All these approaches have their own advantages and disadvantages, and are sources of specific artifacts. Special care should be taken when positioning cameras or when selecting rendering parameters. Unnatural correspondences between the images in a stereo-pair (i.e. vertical disparity) are source of many types of artifacts<sup>11</sup>. As perfectly parallel camera setup is practically impossible, rectification is an unavoidable pre-processing stage.
- Representation format – Although there are many different formats for encoding 3D video, three main groups have evolved: *multiview video*, where two or more video streams show the same scene from different viewpoints; *video-plus-depth*, where each pixel is augmented with information of its distance from the camera; and *dynamic 3D meshes*, where 3D video is represented by dynamic 3D surface geometry<sup>20</sup>. Video-plus-depth format is suitable for multiview displays, as it can be used regardless of the number of views a particular screen provides<sup>20</sup>. On the

downside, video-plus-depth rendering requires interpolation of occluded areas, which causes *disocclusion artifacts*. This is addressed by using layered depth images (LDI), or multiview video-plus-depth encoding<sup>21</sup>. If the representation format is different from the one the scene was originally captured, format conversion of another source of artifacts. Some artifacts which are common in one format and not possible in another – for example in video-plus-depth disocclusion artifacts are common, while vertical parallax does not occur.

- Coding – there are various coding schemes, which utilize temporal, spatial or inter-channel similarities of a 3D video<sup>22</sup>. Two approaches are most popular for stereo-video – multi-view coding standardized as an amendment to H.264/AVC<sup>1, 2</sup>, and dense depth representation where depth channel can be compressed using H.264/AVC and stored in MPEG container<sup>23, 24</sup>. Special care should be taken when algorithms originally designed for single video channel are used for stereoscopic video, as important binocular depth-cues may be lost.
- Transmission – in the case of digital wireless transmission a common problem is burst packet losses<sup>25</sup>. Resilience and error concealment algorithms attempt to mitigate the impact on the video, but if not designed for stereo-video, such algorithms might introduce additional artifacts on their own.
- Visualization – there are various approaches for 3D scene visualization, offering different degree scene approximation<sup>26</sup>. Each family of 3D displays has its own characteristic artifacts, and the artifacts are often scene dependant<sup>7, 11</sup>.

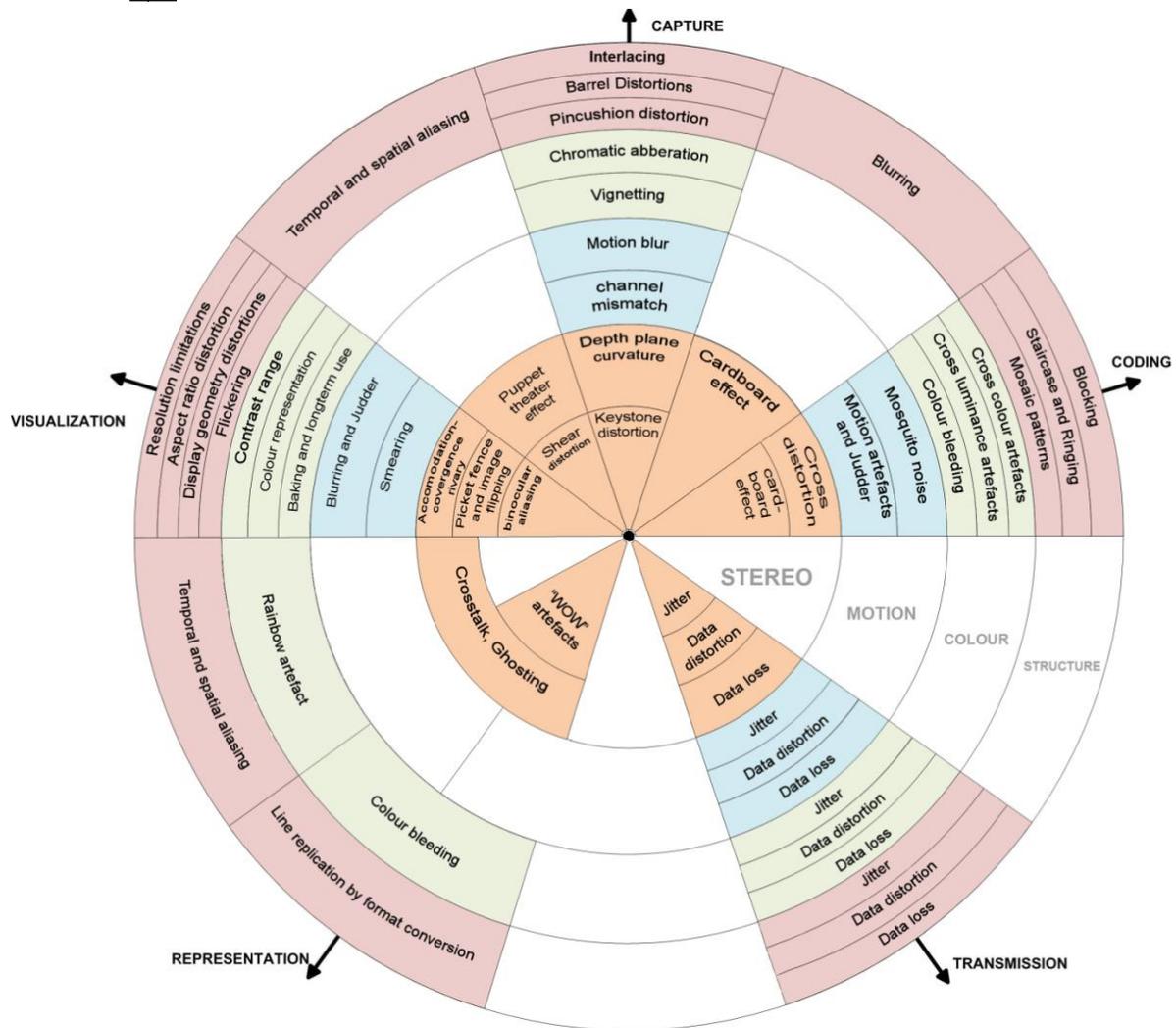


Figure 3, Artifacts, caused by various stages of content delivery and affecting various “layers” of human depth perception.

As a result, stereoscopic artifacts might be created during various stages in the mobile 3DTV content delivery, and might affect different “layers” of human 3D vision, as shown in [Figure 3](#).

#### 4. ARTIFACT CLASSIFICATION

In 3D video, many causes might lead to unnatural scene representations. For building taxonomy of stereoscopic artifacts, we use a top-down approach: first we identify content delivery stages, which might create artifacts, and then we speculate if and how these artifacts will affect various stages of human perception of depth. Our classification is presented in [Table 1](#). The columns represent the causes for artifacts, coming from different content delivery stages – capture, representation, coding, transmission and visualization. The rows are groups of artifacts as they are interpreted by the “layers” of human vision – structure, color, motion and binocular. These layers roughly represent the visual pathways as they appeared during the successive stages of evolution. By *structure* we denote the spatial (and color-less) vision. It is assumed that during the evolution human vision adapted for assessing the “structure” (contours and texture) of images<sup>27</sup>, and some artifacts manifest themselves as affecting image structure. *Color* and *motion* rows represent the color and motion vision, accordingly. As we noted before, all artifacts in the table affect the binocular depth perception. However, the row designated with *binocular* contains artifacts which have meaning only when perceived as a stereo-pair. In other words, these are artifacts that cannot be perceived with a single eye (e.g. vertical disparity).

Table 1 – Classification of stereoscopic artifacts

	Capture	Representation/ Conversion	Coding	Transmission / Error Resilience	Visualization
STRUCTURE	<ul style="list-style-type: none"> <li>- blur by defocusing</li> <li>- barrel distortions</li> <li>- pincushion distortions</li> <li>- interlacing</li> <li>- temporal and spectral aliasing</li> <li>- downsampling</li> <li>- noise introduction</li> </ul>	<ul style="list-style-type: none"> <li>- temporal and spatial aliasing</li> <li>- line replication</li> </ul>	<ul style="list-style-type: none"> <li>- blocking artefacts</li> <li>- mosaic patterns</li> <li>- staircase effect</li> <li>- ringing</li> </ul>	<ul style="list-style-type: none"> <li>- data loss</li> <li>- data distortion</li> <li>- jitter</li> </ul>	<ul style="list-style-type: none"> <li>- flickering</li> <li>- resolution limitations</li> <li>- aspect ratio distortions</li> <li>- display geometry distortions</li> <li>- spatial aliasing by subsampling on non-rectangular grid)</li> </ul>
COLOR	<ul style="list-style-type: none"> <li>- Chromatic aberration</li> <li>- Vignetting- decreasing intensity</li> </ul>	<ul style="list-style-type: none"> <li>- temporal and spatial aliasing</li> </ul>	<ul style="list-style-type: none"> <li>- cross- colour artefacts</li> <li>- colour bleeding</li> </ul>	<ul style="list-style-type: none"> <li>- color bleeding</li> </ul>	<ul style="list-style-type: none"> <li>- contrast range</li> <li>- colour representation</li> <li>- baking and longterm use</li> <li>- viewing angle dependant colour representation</li> <li>- rainbow artefact</li> </ul>
MOTION	<ul style="list-style-type: none"> <li>- motion blur</li> <li>- temporal mismatch</li> </ul>		<ul style="list-style-type: none"> <li>- motion compensation artefacts</li> <li>- mosquito noise</li> <li>- Judder</li> </ul>	<ul style="list-style-type: none"> <li>- loss/distortion in motion</li> <li>- jitter</li> </ul>	<ul style="list-style-type: none"> <li>- smearing</li> <li>- blurring and judder</li> </ul>
BINOCULAR	<ul style="list-style-type: none"> <li>- depth plane curvature</li> <li>- keystone-distortion</li> <li>- cardboard effect</li> </ul>	<ul style="list-style-type: none"> <li>- ghosting by disocclusion</li> <li>- Perspective-binocular rivalry ("WOW"- artefacts)</li> </ul>	<ul style="list-style-type: none"> <li>- cross distortions</li> <li>- cardboard effect</li> <li>- depth "bleeding"/depth "ringing"</li> </ul>	<ul style="list-style-type: none"> <li>- data loss, one channel</li> <li>- data loss, propagating</li> </ul>	<ul style="list-style-type: none"> <li>- shear distortion</li> <li>- ghosting by crosstalk</li> <li>- angle dependant binocular aliasing</li> <li>- accomodation convergence rivalry</li> <li>- lattice artefacts</li> <li>- puppet theater effect</li> <li>- picket fence effect</li> <li>- Image flipping (Pseudoscopic Image)</li> </ul>

The process of artifact mapping is not always straightforward – sometime one stage in the dataflow might cause several types of artifacts, while sometimes artifacts created in different stages are perceived in a similar way (e.g. ghosting). Following that, the diagram from [Figure 3](#) cannot be easily translated to a flat table. Some artifacts are listed repeatedly, while some artifacts groups span across multiple cells. Furthermore, some combinations of rows (causes for artifacts) and columns (artifact manifestations) are omitted as unrelated to the usage scenario of mobile 3DTV.

#### 5. ARTIFACT SIMULATION FRAMEWORK

Not all of the stereoscopic artifacts are likely to affect a mobile 3DTV system. Some of them are not applicable to mobile device due to the technology used (e.g. LCD display, DVB-H transmission). Some of them, cannot be solved through the means of signal processing, and are usually addressed by content providers and/or display manufacturers.

We introduce an artifact simulation channel, which is able to introduce an arbitrary combination of artifacts to a video, with controlled amount of impairment for each artifact. Based on the research from our previous report, artifacts are organized in groups, which follow the flow of a mobile 3D video over a DVB-H channel. Each group of artifacts corresponds to a specific block of our simulation channel as shown in Figure 4. An arbitrary combination of artifacts can be introduced by this channel, but they are always introduced in a certain order – i.e. capture artifacts will always be added before transmission ones.

The first block simulates artifacts caused by sensor limitations. Then, the degraded scene observation is sent to a block which simulates geometric distortions as the ones caused by the camera optics. The next two blocks add global spatial and temporal differences between the video channels, simulating artifacts caused by multi-camera topology and temporal misalignment. The next two blocks simulate spatial and temporal artifacts caused by coding. Then, transmission losses are simulated in the encoded stream. For the case of dense depth video representation, format conversion artifacts are added. Finally, visualization artifacts are added, independent of the position of the observer, or alternatively, for a given observation position.

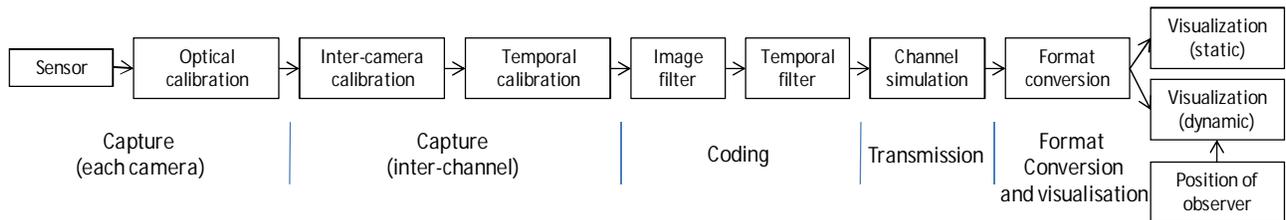


Figure 4, Artifact simulation channel

Following this concept, we have developed a framework for simulation of mobile 3DTV artifacts. The framework is thoroughly described in<sup>30</sup>. It is organized as a collection of Matlab functions, each one responsible for introducing a specific artifact. Additionally, there is a program module, which executes the simulation functions as prescribed by a configuration file. The configuration is stored in a text file, which describes the input and output video streams, the set of artifacts to be introduced, and the parameters for each artifact. One configuration file can specify a set of artifact parameters, which to be applies over several input video files in “batch mode”.

The framework operates on stereo-video streams (where left and right channel are provided as separate video files) or dense depth video streams (where video and depth channel are provided as two separate video files). Video is decoded into a set of frames, each frame is processed and the result is encoded in a video stream again. The blocks of the framework are shown on Figure 5, and are as follows:

- GUI – provides two alternative ways to prepare a configuration file – using Microsoft Excel sheet or using Matlab GUI. The Excel-based GUI uses VBA-scripting. Alternatively, a configuration file can be prepared using a text editor<sup>30</sup>.
- Session manager – opens and parses a set of configuration files.
- General logic – imports video streams or collections of frames; processes them as described in the configuration file; exports video stream or a set of frames.
- Low level processing – introduces artifacts to a given video by processing each frame separately. While applying artifact to one frame, information from other frames or video channels might be used.
- Database of artifact simulation functions – a set of function, each one responsible for introducing a specific artifact. Most of the functions are implemented on Matlab. Three functions – “2D+Z to multiview conversion”, “Multiview to 2D+Z conversion” and “DVB-H packet loss simulation” are implemented as Windows executables, and are called as external functions by the framework.

The next section describes the full list of artifacts, which can be introduced by our framework.

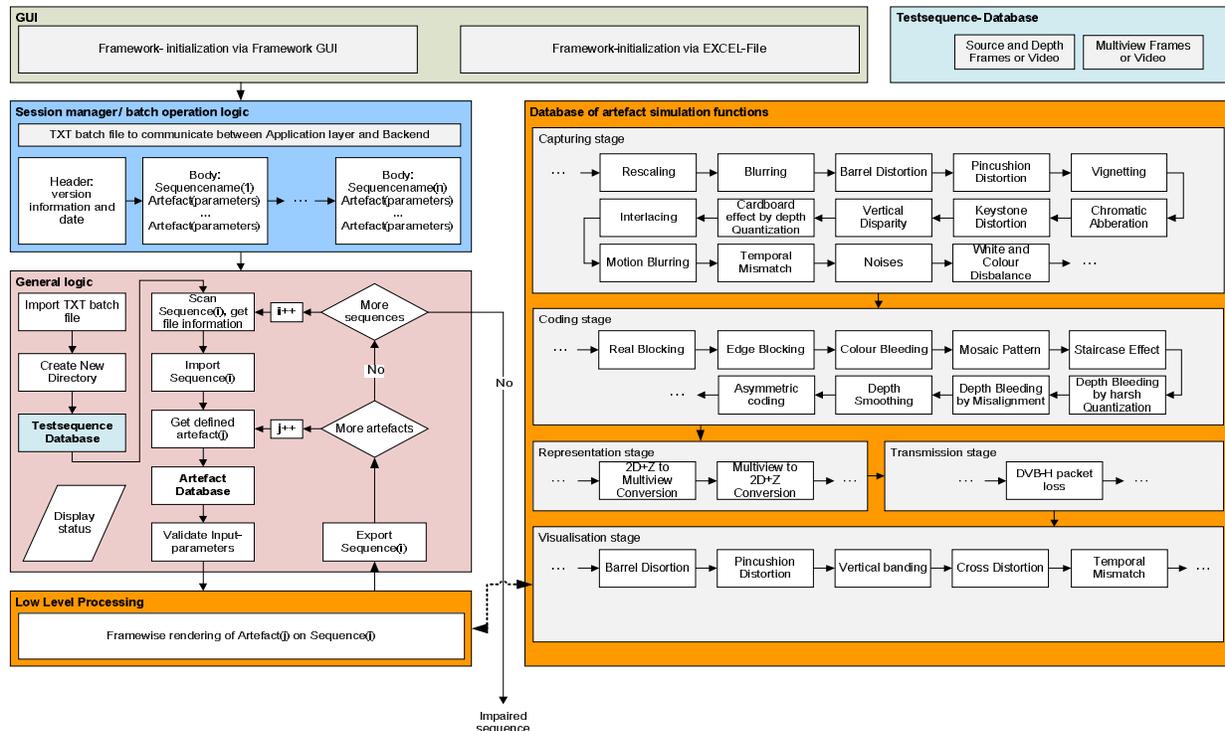


Figure 5, Block diagram of the artifact simulation framework

## 6. MOBILE 3DTV ARTEFACTS SELECTED FOR SIMULATION

### 6.1 Capture artifacts

The capturing process for mobile 3DTV video is similar to the one for a 3DTV system targeting large displays. One thing which separates video broadcast system from video conferencing one is that capture for the former is done off-line and non-real-time, and significant processing power might be spent for producing the best output possible.

We have chosen for simulation the following list of common stereo video capture artifacts:

- *Size and resolution changes* – the problem of choosing the proper resolution for capturing of 3D contents is not necessarily a simple one. Two problems might arise from content resizing – aliasing and wrong disparity range. The perceptual impact of aliasing on stereoscopic video is yet to be studied – if is it going to be masked by binocular suppression, or is going to destroy important texture-based binocular cues. Additionally, changing the size of a multiview 3D video changes the inter-channel relations as well, which might result in a disparity either too small or too large for proper 3D effect. Our framework allows rescaling of video content using various interpolation methods.
- *Blur* might be caused by low-quality optics or wrong focal setting. In a 2D movie, in most cases small amount of blur is permissible. In a binocular setup, predicting how blur will affect quality is more complex task. Depending on the case, blur in one channel might go unnoticed, or in rare cases even improve the perceived quality.
- *Motion blur* – this is usually caused by capturing in low light conditions. The temporal masking and perception of motion blur in stereo video is yet to be studied.
- *Barrel/Pincushion distortion* is a geometrical distortion, which affects each camera separately. In multi-camera it could cause serious artifacts in stereoscopic image, and induce eye-strain. This is corrected by rectification. These artifacts are simulated by applying identical geometric transformation separately to each channel.
- *Keystone distortion* affects the geometric relation between two channels. The result is a trapezoidal shape in opposite direction in left and right camera inputs. It is mainly caused by camera optics and selected multi-camera topology. The presence of keystone distortion can induce eye-strain or fully break the 3D effect of a stereo video. It

also will greatly diminish the precision of dense depth estimation algorithms. Image rectification compensates this effect. Keystone distortion is introduced by simulation of converging camera setup – namely applying projective distortion to each channel, with opposite projection directions.

- *Temporal mismatch* occurs when a 3D scene is shot with multiple cameras, which are not shutter-synchronized. As a result, the frames in both channels are not shot simultaneously, but with slightly shifted in time. While precise time synchronization is of crucial importance for dense depth estimation algorithms, the human visual system can tolerate some amount of time mismatch, without diminishing the perceptual quality. We simulate temporal mismatch by adding temporal offset to one of the channels.
- *Color mismatch* – Some factors (i.e. bright objects with large disparity between cameras) in can cause mismatch in the colors in the images of a scene captured by different cameras. It is most commonly caused by white balance done separately in each camera. Color mismatch is introduced by mimicking automatic white balancing algorithms, with selectable illumination parameters.
- *Interlacing* – Interlaced video is created by scanning the odd and the even lines of an image sensor separately. Interlaced video exhibits specific “jagged-border” artifacts as seen on [Figure 6](#). In 2D video interlacing overlaps consecutive frames in time. As one of the methods for encoding stereo-video involves using of odd and even fields, interlacing might also interleave simultaneous frames from different channels.
- *Cardboard effect* refers to unnatural flattening of objects in stereoscopic images, as if they were cardboard cut-outs<sup>12</sup>. It is believed that the main reason is the field of view of a stereoscopic display being different from the field of view of the scene, thus creating inappropriate depth scaling<sup>28</sup>. In our framework we simulate cardboard effect only on video streams with dense depth. However, our framework could be extended to simulate cardboard effect in other video formats.

Additionally, we are simulating less common capture artifacts like *noise*, *vignetting* and *chromatic aberration*. Proper simulation of *camera noise* is a very demanding task, but usually it is dealt with separately inside each camera. We included noise simulation for the ability to prepare subjective test material where asymmetric amount of noise is present in each channel.



Figure 6, Example of monoscopic 3D artifacts introduced during capture: a) barrel distortion, b) pincushion distortion, c) interlacing and d) color aberration

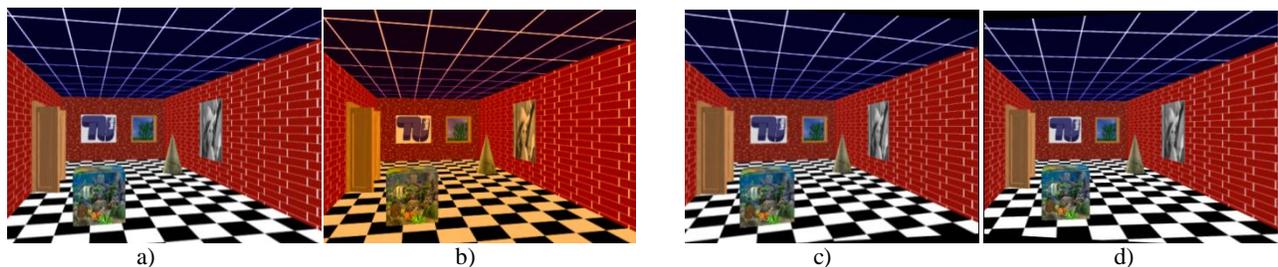


Figure 7, Example of stereoscopic 3D artifacts introduced during capture: a), b) stereoscopic pair exhibition color mismatch, and c), d) stereoscopic pair with added keystone distortion

## 6.2 Coding artifacts

While the visibility of coding artifacts is quite well studied for 2D case, the impact on the 3D vision is yet to be determined.

- Transform-caused artifacts come from the transforms and quantization used for compressing the video stream. *Blocking, mosaic patterns, staircase effect, ringing, and color bleeding* artifacts are in this group. All of them are well visible, and as overlay structural changes on the image, they might destroy depth cues and even create misleading ones. *Depth bleeding* and *depth ringing* are artifacts specific for the coding the depth map of a scene, and as such, they exist only in dense depth-based 3D video representations. Notably, such artifacts can be mitigated by using structural information of the 2D scene.
- Temporal coding artifacts appear as a result of transform/quantization over time. Temporal inconsistency such as *mosquito* noise is the most common artifact in this group. Artifacts caused by imprecise *motion prediction* are also possible. This group of artifacts can appear both in multi-view and in dense depth 3D video.

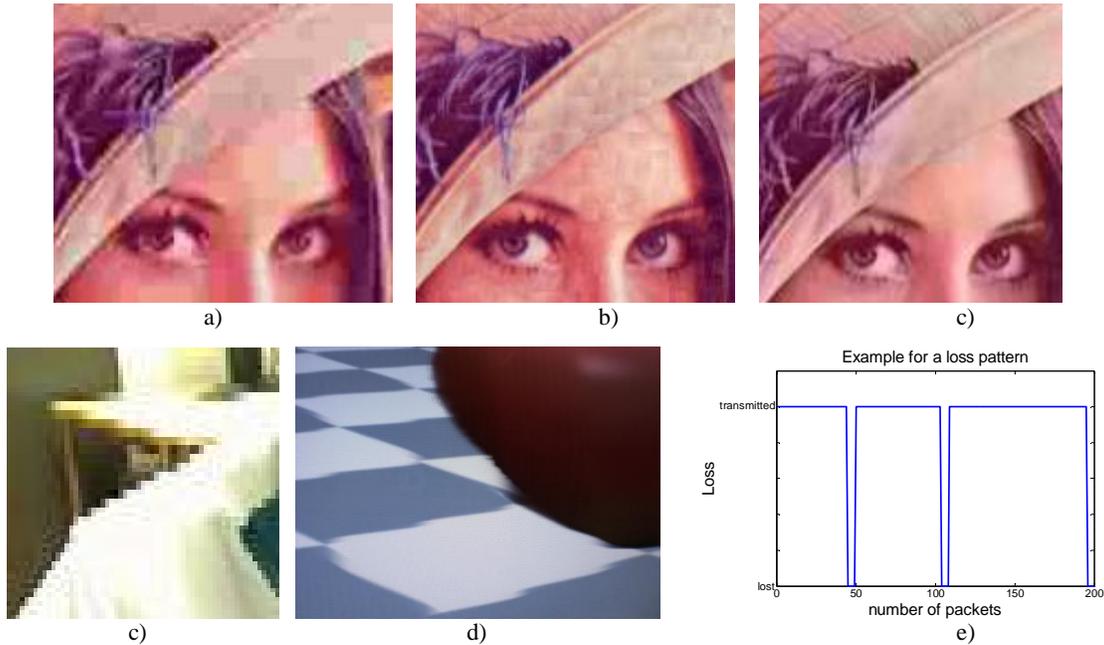


Figure 8, a)-d) Example of coding artifacts: a) blocking by harsh quantization, b) blocking by edge discontinuities, c) color bleeding, d) staircase effect, depth bleeding; e) example error pattern of DVB-H channel losses.

Our framework simulates the following coding artifacts:

- *Blocking by harsh quantization* is among the most widely studied distortions in video coding. The most common source of the artifact is block-based DCT compression, which involves quantization and entropy coding of the results. This process creates a number of image impairments, the most noticeable of which is discontinuities at the boundaries of the encoded blocks. In our framework, we simulate blocking by harsh quantization by utilizing the DCT block-based compression used in JPEG. The results are seen on Figure 8a. Additionally, some authors propose that blocking might be considered as several, visually separate artifacts – *block-edge discontinuities, color bleeding, blur* and *staircase artifacts*<sup>18</sup>. Our artifact also provides means to simulate these artifacts separately, if needed.
- *Block-edge discontinuities* – block-based coding tries to exploit the spatial correlation between pixels in a picture, but does not take into account the possible correlation beyond the block borders. One important property of such distortion is that the mean intensity of the block remains the same as before. In our framework, we provide an option that block-edge discontinuities are simulated separately from block-based DCT artifacts. We simulate block-edge discontinuities by introducing luminance distortions inside of a block while keeping the mean luminance constant, as seen in Figure 8b.

- *Color bleeding* is an artifact caused by harsh quantization of high frequency chrominance coefficients. Since chrominance is typically sub-sampled, bleeding can occur beyond the range of a block. Color bleeding is simulated by applying different levels of quantization to chrominance and luminance channels, as seen in [Figure 8c](#).
- *Staircase effect* affects diagonal edges of a picture. The quantization of DCT coefficients results in diagonal lines which are almost horizontal or almost vertical, to be represented a series of blocks. We approximate the effect of staircase artifact by selective pixel doubling in horizontal or vertical direction, which produces staircase edges as the ones seen in [Figure 8c](#).
- *Cross-distortion* is an artifact caused by asymmetrical stereo-video coding. The asymmetry might be both in spatial (one channel with lower resolution) or in temporal (one channel having lower frame-rate) domains. The effect of spatial or temporal sub-sampling of one channel is not yet thoroughly studied. Asymmetrical coding is applied for multi-view video only.

Additionally, our framework simulates less common coding artifacts which affect videos in image plus depth format – *depth bleeding* and *depth smoothing*. Depth bleeding is caused by a process similar to the one, which causes color bleeding – with the difference that it degrades the depth channel instead of the chrominance. Depth smoothing is could be caused by asymmetric compression or resolution of the depth channel. In some cases, depth smoothing might improve the quality of an image plus depth video, as it will hide some disocclusion artifacts.

### 6.3 Conversion artifacts

Format conversion artifacts occur during the conversion for a dense-depth representation used for broadcast to a multiview one as needed by the display. Most common here are *disocclusion artifacts*, which are more pronounced when rendering observations at angles much different from the central observation point, and less pronounced when layered depth images are used<sup>21</sup>. *Perspective-stereopsis rivalry* occurs if the conversion over-exaggerates the depth levels in the depth map. *Temporal inconsistency* of the depth estimation creates artifacts similar to mosquito and depth ringing. It is quite difficult to simulate conversion artifacts separately from the actual process of conversion. Our framework allows various types of conversion algorithms and quality settings to be used for introducing of conversion artifacts.

### 6.4 Transmission artifacts

The presence of artifacts generated in the transmission stage depends very much on the coding algorithms used and how the decoder copes with the channel errors. In DVB-H transmission most common are burst errors<sup>29</sup>, which results in packet losses distributed in tight groups. In MPEG-4 based encoders packet losses might result in propagating or non-propagating errors, depending on where the error occurs in respect to the I-frames, and the ratio between I- and P-frames. We simulate transmission errors by obtaining error patterns of the DVB-H channel and use them for simulation of channel losses as it is done in<sup>29</sup>. Example DVB-H error pattern is shown in [Figure 8e](#).

### 6.5 Visualization artifacts

Artifacts in visualization of mobile 3DTV are caused by limitations of the display technology used. We expect a mobile 3DTV system to use 2-view, autostereoscopic display. Such displays use spatial multiplexing of the channels, and the visibility of all artifacts depends on the position of the observer.

Some visualization artifacts are perceived while changing the position in respect to the display. Such artifacts are *angle dependant color representation*, *pseudostereoscopy*, *picket fence effect*, or the unnatural image parallax causing *shear distortion*. Others appear only for some observation angles, as *image flipping*. The artifacts in this group are difficult to simulate, but easy to mitigate for a given position of the observer.

In our framework, we simulate only artifacts, visible by a static observer:

- *Vertical banding* can be regarded as the “static” version of picket fence effect. It is very common for displays with parallax barrier, and manifests itself as changes of the intensity across the display – as if dark vertical bands are superimposed on the image. Even though it depends on the viewing angle, it is visible from most of the viewing angle/observation distance combinations, except for a few observation “sweet spots”. An example of simulated vertical banding can be seen in [Figure 9a](#).
- *Temporal mismatch* is a temporal misalignment between the video channels. While during capture such misalignment is usually very small, but depending on the decoder, temporal misalignment can increase to to several seconds. Typical causes are reception problems and rudimentary error concealment.

- *Resolution change*– it is possible that a stereo-video stream needs to be rescaled on the receiving device. Rescaling can create the same problems as doing the same during capture – aliasing and improper disparity. Additionally, rescaling during visualization might affect (exaggerate or suppress) other artifacts.
- *Cross-talk* – display imperfections can cause cross-talk and other forms of inter-channel distortion. Stereo- and multiview displays using parallax barrier are particularly vulnerable to crosstalk. Crosstalk is simulated by scaling the intensity of a frame in one channel and superimposing it over a frame of the other channel, as seen in Figure 9b.

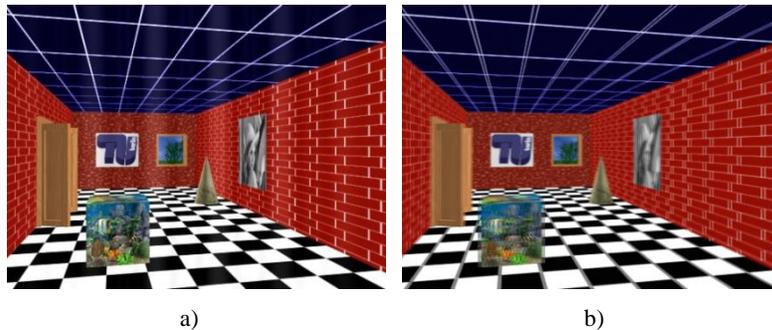


Figure 9, Example of 3D artifacts introduced in the visualization stage: a) banding artifacts, and b) crosstalk.

## 7. CONCLUSION

We identified which 3D artifacts can occur in a mobile 3DTV system, featuring H.264 AVC type of encoding, DVB-H transmission channel and portable autostereoscopic display. We discussed how different stages of mobile 3DTV content delivery could affect the subsystems of human 3D vision. We proposed artifact simulation channel which follows the natural flow of a mobile 3D video over a DVB-H channel.

We presented an artifact simulation framework that allows an arbitrary combination of artifacts to be introduced to 3D video. Such framework can be used to perform subjective experiments, in which the perceptual quality of various mobile 3DTV artifacts can be estimated.

## 8. ACKNOWLEDGEMENT

This work is supported by the European Commission within the ICT Programme of FP7 under Grant 216503 with the acronym MOBILE3DTV.

## REFERENCES

- <sup>1</sup> ISO/IEC JTC1/SC29/WG11, “Study Text of ISO/IEC 14496-10:2008/FPDAM 1 Multiview Video Coding”, Doc. N9760, Archamps, France, May 2008.
- <sup>2</sup> ISO/IEC JTC1/SC29/WG11, “Joint Multiview Video Model 8”, Doc. N9762, Archamps, France, May 2008.
- <sup>3</sup> Z. Tan and A. Zakhor, “Error control for video multicast using hierarchical FEC,” in *Proc. of the Int. Conf. on Image Processing*, Kobe, Japan, October 1999, vol. 1, pp. 401-405.
- <sup>4</sup> G. J. Woodgate, J. Harrold, “Autostereoscopic display technology for mobile 3DTV applications”, in *Proc. SPIE* Vol.6490A-19, 2007
- <sup>5</sup> Sharp Laboratories of Europe, website, [http://www.sle.sharp.co.uk/research/optical\\_imaging/3d\\_research.php](http://www.sle.sharp.co.uk/research/optical_imaging/3d_research.php)
- <sup>6</sup> S.Uehara, T.Hiroya, H. Kusanagi; K. Shigemura, H.Asada, “1-inch diagonal transreflective 2D and 3D LCD with HDDP arrangement”, in *Proc. SPIE-IS&T Electronic Imaging 2008, Stereoscopic Displays and Applications XIX*, Vol. 6803, San Jose, USA, January 2008
- <sup>7</sup> S. Jumisko-Pyykkö and J. Häkkinen, “Evaluation of subjective video quality of mobile devices”, in *MULTIMEDIA '05: Proc. 13th ACM international conf. on Multimedia*. New York, NY, USA: ACM Press, pp. 535–538, 2005.
- <sup>8</sup> J. Häkkinen, M. Liinasuo, J. Takatalo, and G. Nyman, “Visual comfort with mobile stereoscopic gaming”, *Proceedings of SPIE*, vol. 6055, p. 60550A, 2006.

- <sup>9</sup> M. McCauley and T. Sharkey, "Cybersickness: Perception of Self-Motion in Virtual Environments" in *Presence: Teleoperators and Virtual Environments*, 1(3), 311-318., 1992
- <sup>10</sup> L. Meesters, W. IJsselsteijn, P. Seuntiëns, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, No. 3, 2004, pp. 381 – 391.
- <sup>11</sup> W. IJsselsteijn, P. Seuntiëns and L. Meesters, "Human factors of 3D displays", in (Schreer, Kauff, Sikora, eds.) *3D Video Communication*, Wiley, 2005.
- <sup>12</sup> Wandell, B.A., *Foundations of vision*, Sinauer Associates, Inc, Sunderland, Massachusetts, USA, 1995.
- <sup>13</sup> D. Chandler, "Visual Perception (Introductory Notes for Media Theory Students)", MSC portal site, University of Wales, Aberystwyth, available at <http://www.aber.ac.uk/media/sections/image05.html>
- <sup>14</sup> IP. Howard and BJ Rogers, "Binocular Vision and Stereopsis", in Oxford Univ. Press, NY, Oxford, 1995
- <sup>15</sup> M. Wexler and J. Boxtel, "Depth perception by the active observer", *Trends in Cognitive Sciences*, 9, pp. 431-438, Sept, 2005
- <sup>16</sup> Julesz, B. *Foundations of Cyclopean Perception*, The University of Chicago Press, Chicago, 1971.
- <sup>17</sup> A.Boev, A. Gotchev, K. Egiazarian, A. Aksay and G. Akar, "Towards compound stereo-video quality metric: a specific encoder-based framework". *Proc. of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI 2006)*, Denver, CO, USA, 2006.
- <sup>18</sup> M. Yuen, "Coding Artifacts and Visual Distortions", in (H. Wu. K. Rao eds), *Digital Video Image Quality and Perceptual Coding*, ISBN 9780824727772 , CRC Press, 2005
- <sup>19</sup> Merriam-Webster's online dictionary, available at <http://www.merriam-webster.com/dictionary/artifact>
- <sup>20</sup> A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, A.Koz, "Coding Algorithms for 3DTV—A Survey," *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.17, no.11, pp.1606-1621, Nov. 2007
- <sup>21</sup> A. Alatan, Y. Yemez, U. Gudukbay, X. Zabulis, K. Muller, C. Erdem, C. Weigel, A., "Scene Representation Technologies for 3DTV—A Survey," *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.17, no.11, pp.1587-1605, Nov. 2007
- <sup>22</sup> A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, A. Koz, "Coding Algorithms for 3DTV—A Survey," *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.17, no.11, pp.1606-1621, Nov. 2007
- <sup>23</sup> ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information", Doc. N8768, Marrakech, Morocco, January 2007.
- <sup>24</sup> ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 13818-1:2003/FDAM2 Carriage of Auxiliary Data", Doc. N8799, Marrakech, Morocco, January 2007.
- <sup>25</sup> Lin, C., Ke, C., Shieh, C., and Chilamkurti, N. K. 2006. The Packet Loss Effect on MPEG Video Transmission in Wireless Networks. In *Proc. 20th international Conference on Advanced information Networking and Applications - Volume I (Aina'06) - Volume 01* (April 18 - 20, 2006). AINA. IEEE Computer Society, Washington, DC, 565-572.
- <sup>26</sup> P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, 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
- <sup>27</sup> Z. Wang, , A. Bovik, H. Sheikh and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Trans. Image Processing*, vol. 13, No. 4, 2004, pp. 600-612
- <sup>28</sup> I. Howard and B. Rogers, *Binocular Vision and stereopsis*, Oxford University press, New York, 1995
- <sup>29</sup> J. Poikonen, J. Paavola, "Error Models for the Transport Stream Packet Channel in the DVB-H Link Layer", *Proc. ICC 2006*, Istanbul, Turkey, 2006
- <sup>30</sup> A. Boev, D. Hollosi, and A. Gotchev, "Software for simulation of artifacts and database of impaired videos", Mobile3DTV Project report, available on <http://mobile3dtv.eu>