



**HAL**  
open science

## Towards a New Quality Metric for 3-D Synthesized View Assessment

Emilie Bosc, Romuald P epion, Patrick Le Callet, Martin K oppel, Patrick Ndjiki-Nya, Muriel Pressigout, Luce Morin

► **To cite this version:**

Emilie Bosc, Romuald P epion, Patrick Le Callet, Martin K oppel, Patrick Ndjiki-Nya, et al.. Towards a New Quality Metric for 3-D Synthesized View Assessment. *IEEE Journal of Selected Topics in Signal Processing*, 2011, pp.J-STSP-ETVC-00048-2011. hal-00628070

**HAL Id: hal-00628070**

**<https://hal.science/hal-00628070>**

Submitted on 30 Sep 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche franais ou  trangers, des laboratoires publics ou priv es.

# Towards a New Quality Metric for 3-D Synthesized View Assessment

Emilie Bosc, *Student Member, IEEE*, Romuald P epion, Patrick Le Callet, *Member, IEEE*,  
Martin K oppel, Patrick Ndjiki-Nya, Muriel Pressigout, and Luce Morin

**Abstract**—3DTV technology has brought out new challenges such as the question of synthesized view evaluation. Synthesized views are generated through a depth image-based rendering (DIBR) process. This process induces new types of artifacts whose impact on visual quality has to be identified considering various contexts of use. While visual quality assessment has been the subject of many studies in the last twenty years, there are still some unanswered questions regarding new technological improvement. DIBR brings new challenges mainly because it deals with geometric distortions. This paper considers the DIBR-based synthesized view evaluation problem. Different experiments have been carried out. They question the protocols of subjective assessment and the reliability of the objective quality metrics in the context of 3DTV, in these specific conditions (DIBR-based synthesized views). They consist in assessing seven different view synthesis algorithms through subjective and objective measurements. Results show that usual metrics are not sufficient for assessing 3-D synthesized views, since they do not correctly render human judgment. Synthesized views contain specific artifacts located around the disoccluded areas, but usual metrics seem to be unable to express the degree of annoyance perceived in the whole image. The study provides hints for a new objective measure. Two approaches are proposed: the first one is based on the analysis of the shifts of the contours of the synthesized view; the second one is based on the computation of a mean SSIM score of the disoccluded areas.

**Index Terms**—Depth image-based rendering (DIBR), multi-view, quality assessment, quality metrics, video plus depth, view synthesis.

THE emerging 3-D video applications have encouraged numerous investigations in various fields from video acquisition to display technologies. These applications are known as 3-D television (3DTV) and free viewpoint video (FVV) [1]. 3DTV provides a depth feeling thanks to an appropriate 3-D display. FVV interactively allows the user to control the viewpoint in the scene. Considering the demand for high-quality visual content, the success of 3-D video applications is closely related to its ability to provide viewers with a high quality level of visual experience.

A reliable assessment protocol for synthesized views is urgently needed. The brand-new 3-D applications bring out new elements but, so far, their complexity may have been under-estimated. Thanks to recent studies that brought this complexity into focus, it is now more understandable. Sentiens [2] reported that not only image quality should be taken into account for an assessment framework, but visual comfort and depth feeling as well. These last aspects are still under investigation ([3], [4]). Concerning image quality, new distortion types have appeared, as discussed in [5], [6], [7]. Some of them are related to the synthesis process, others are related to compression methods or display technologies. A few

of them are:

- *keystone* effect : the image looks like a trapezoid;
- *ghosting* effect: this is a shadow-like artifact;
- *cardboard* effect: depth is perceived as unnatural, as discrete incoherent planes;
- *puppet-theater* effect: characters in the scene appear miniaturized;
- *staircase* effect: discontinuities appear between adjacent blocks.

El-Yamany et al. [8] showed that distortions from compression may be masked by distortion from the synthesis process. Consequently, they recommend using the control synthesis (i.e., image synthesized from uncompressed data) as a reference when assessing codec performances. Concerning the subjective evaluation sessions, another aspect to be considered is the video content: as mentioned in [9], compared to 2-D, video content may be determinant in the evaluation by observers.

Consequently, considering the factors brought by 3-D video, both subjective and objective methods of evaluation have been studied for an adaptation to 3-D. Subjective quality assessment is delicate while addressing a new type of conditions because one has to define the optimal way to get reliable data. In [10], the authors propose a strategy to develop a quantitative model that estimates 3-D perceptual quality, in the context of mesh representation of 3-D objects. Based on a prior study on quality factors such as geometry resolution, texture resolution, and shading complexity, the authors derived a model that approximates the perceptual quality of mesh-based 3-D objects. In our study, i.e. in the case of use of Multiview video plus depth (MVD) data and DIBR algorithms, the right protocol to assess the visual quality with observers is still an open question, since new conditions are introduced. The adequate protocol might vary according to the purpose ( impact of compression, of DIBR techniques comparison). In the absence of any better subjective quality assessment methodologies for the evaluation of synthesized views, the measurement is mostly obtained through 2-D validated protocols. Chen et al. [11] proposed new requirements for subjective video quality assessment. In particular, the authors point out the fact that the minimum of 15 observers, as recommended in ITU-BT.500, may not be sufficient because of the fluctuation of viewers' assessment in 3-D. Therefore, they suggest reconsidering the statistical analysis of viewers' opinion for appropriate rejection of incoherent responses. For 3DTV applications, the authors essentially explained how test material (3-D displays and content disparity) affects depth rendering. This calls for

a careful analysis of future experiment results. Concerning objective methods, various approaches have been proposed but it is still a delicate topic. In the context of polygonal meshes based representation, in [12], the authors introduce the concept of just-noticeable-difference for meshes: they investigate a perceptual threshold where humans can just distinguish the difference between two levels-of-details (this refers to the complexity of the 3-D mesh).

The added value, compared to 2-D conventional video, comes from the exploitation of multiple video sequences acquired at different viewpoints in the scene. These video sequences can be processed into different 3-D representations [13]: image-based representations (conventional stereoscopic video, multi-view video, etc.), surface-based representations (polygonal meshes, etc.), point-based representations or depth image-based representations (2-D+Z, multi-view video plus depth, layered depth video, etc.) among others. This study is in line with the depth image-based representation context, especially using multi-view video plus depth data, referred to as MVD in the rest of the paper.

MVD designates the association of multiple conventional color videos, referred to as “texture data,” and their correspondent depth video sequences, referred to as “depth data.” Depth image-based rendering (DIBR) algorithms synthesize a novel view of the scene, different from those captured by the cameras. They require information on scene geometry and texture information of the scene at slightly different viewpoints, provided by depth data and texture respectively.

The problem of assessing 3-D video quality first refers to the object under test. There is no standardized quality assessment framework for 3-D video, so the object under test is chosen depending on the desired application, on the used 3-D representation (mentioned above) and on the chosen display (i.e., stereoscopic, auto-stereoscopic or multi-autoscopic). For instance, in the case of stereoscopic videos, quality of right and left views may be separately assessed, or tools allowing assessing the fused views quality may be used as in [14].

In our application context, the very first criterion to consider is the quality of the synthesized view. It is justified by the following reasons. Firstly, 3DTV technology relies on the stereopsis phenomenon. This designates the fact that the human brain fuses two slightly different images, presented on each eye, and interprets the 3-D content [15]. As a result, 3-D displays should provide the appropriate stereoscopic images to ensure depth feeling. Yet, the captured views may not be stereo-compliant depending on the display characteristics. In that case, view synthesis is needed to create the correct stereoscopic pairs. Secondly, in the case of FVV, smooth navigation into the scene requires the generation of non-acquired views. Thirdly, for broadcast situations, constraints on bandwidth limit the amount of data to be transmitted. Generally, virtual views are synthesized from compressed texture and depth data. So compression performances can be evaluated by the rendered views quality. Fourthly, new compression methods using view synthesis prediction have recently been proposed: Na et al. [16] proposed a coding method for MVD using virtual view synthesis prediction. Morvan et al. [17] proposed a joint depth/texture bit allocation

algorithm for MVD compression that minimizes the distortion of the intermediate rendered view where original view is taken as a reference for the distortion measure. Consequently, our study focuses on the quality of the virtual synthesized view. More precisely, our study considers virtual synthesized views in specific conditions: as a preliminary experiment, and because we would like to free ourselves from the displays whose technology and artifacts are not under our control, we propose to primarily estimate the quality of DIBR-based synthesized views in 2-D. This is actually a plausible context in a FVV application too.

Moving Pictures Expert Group (MPEG, ISO/IEC/JTC 1/SC 29/WG 11) created a working group in order to develop a basic framework that, from a limited number of views, enables FVV. Four exploration experiments (EE) have been established ([18]): EE1 for improved depth generation, EE2 for extrapolation capabilities of view synthesis, EE4 for coding experiments and EE5 for 4-view configurations. They provide specific requirements concerning the test conditions, including coding conditions and objective quality evaluation conditions. Especially, the performances of the proposed tools will be based on Peak Signal-to-Noise ratio (PSNR) in some cases. In the meantime, the Video Quality Expert Group (VQEG) created the 3DTV Work Group ([19]) in order to conduct tests and report the results to ITU or other standards organizations. VQEG is concerned by questions related to the experimental design of subjective assessment methods.

Thus, the quality of different 3-D video formats and 3-D video coding strategies are to be evaluated by researchers, and the outcome will determine the chosen solutions in the industry. Therefore it seems to be of the utmost importance and urgency to confirm that usual quality assessment tools are easily applicable to 3-D synthesized views. Then, in this context, we consider the reliability of the commonly used 2-D video metrics when assessing DIBR rendered virtual views. The contribution of this paper comes from the two main questions we studied. First, how well-adapted are the used subjective assessment protocols in the case of DIBR rendered virtual views? Second, is there a correlation between commonly used 2-D video metrics scores and subjective scores when evaluating the quality of DIBR rendered virtual views? Because DIBR rendered virtual views are 2-D images, usual 2-D metrics are naturally thought of as appropriate for assessing their quality. The problem we consider has not been treated in the conditions we propose, in the literature, yet. In our context, the results of our study confirm the common idea that 3-D video is not just the extension of 2-D video and that commonly used 2-D video metrics are not sufficient to assess the synthesized view quality. The synthesized images quality is measured through pixel-based or perceptual-like metrics. The images are synthesized with seven different DIBR algorithms. Correlation with human perception is estimated through subjective assessments.

This document is organized as follows: Section I is a review of the latest works in the field of quality evaluation in 3-D Video. Section II addresses the view synthesis process. Section III considers the first question regarding the subjective assessment protocols while Section IV studies the reliability

of the objective tools. Section V proposes directions for future work and Section VI concludes the paper.

## I. QUALITY EVALUATION IN THE CONTEXT OF MVD

Previous studies related to investigations on quality metrics for 3-D images are mentioned in this section.

In our context (MVD representation of 3-D scenes), conventional 2-D objective quality models have been spontaneously thought of for assessing 3-D videos, because their efficiency has been proved for the evaluation of 2-D image quality. You et al. [20] addressed their capabilities in the context of stereoscopic images. The authors applied different 2-D quality metrics on a single distorted view. Various types of distortions were studied. They shown that, if used directly, the tested metrics were not robust to changes of distortion types, but if the disparity quality was appropriately combined, their capabilities could be improved.

Previous works also questioned the reliability of commonly used 2-D metrics. Yasakethu et al. [21] studied the relationship between subjective and usual objective measures (PSNR, SSIM, and VQM), in the context of stereoscopic videos and also 2-D+Z videos. Sequences are compressed at different bit-rates and then evaluated by objective measures and through a stereoscopic display and an auto-stereoscopic display with 15 expert observers. Objective metrics are applied to decoded frames, but not to synthesized views as in our study. Yet, an auto-stereoscopic display is used for the subjective measurement, so a synthesis process is used. However, this process used by the display is blind. Statistical tools are used to show that 2-D metrics might be used, separately on each view/type of sequence to assess 3-D video quality. Tikanmaki et al. [22] studied the assessment of 3-D encoded video in the context of 2-D+Z, and the authors considered the synthesized view quality. When trying to determine the optimal ratio between color and depth data, the authors observed that PSNR and VSSIM led to different conclusions, according to the compression choices. PSNR seemed unstable depending on the direction of the targeted virtual viewpoint. Yet, the subjective scores correlated the VSSIM scores.

Concepts of psychophysics, used in 2-D conditions (contrast sensitivity [23], contrast masking and facilitation effects [24], local contrast [25] etc.) can certainly be exploited in these new conditions, but previous studies suggested that new concepts have also been introduced. Our study evaluates the reliability of commonly used 2-D objective and subjective measurements when assessing DIBR-based rendered virtual views. This is of great interest because aforementioned working groups are currently planning to rely on the results of objective and subjective measurements inspired from 2-D protocols, to assess new methods enabling 3-DTV.

## II. VIRTUAL VIEW SYNTHESIS ALGORITHMS

In this section, different DIBR methods are presented. DIBR is the process of synthesizing “virtual” views at a slightly different viewing perspective, using an image or video and the associated per-pixel depth information. This novel view generation can be divided into two steps. In a first step, the

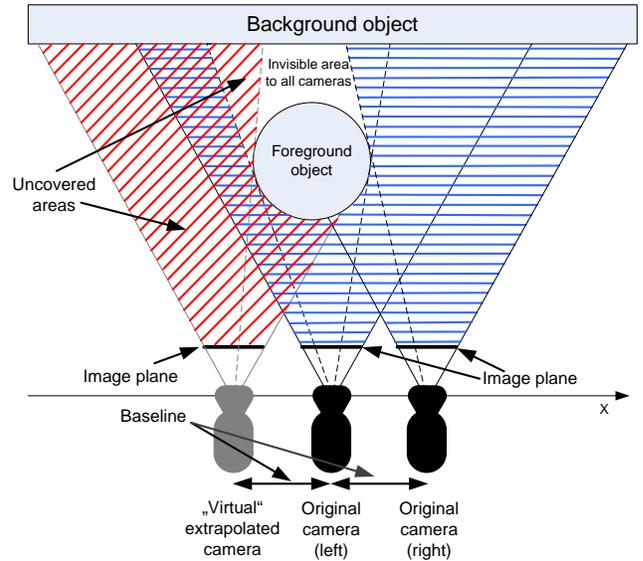


Fig. 1. Extrapolation scenario of a “virtual” camera view from original cameras: Scene areas that are visible in the original cameras are marked with blue horizontal lines. Those visible in the “virtual” view are highlighted with red oblique lines. Uncovered areas are visible in the “virtual” camera but invisible in the original ones.

texture information is re-projected into the 3-D world using the per-pixel depth information. At the next step, these 3-D space points are projected into the 2-D image plane of the “virtual” camera. The re-projection from 2-D to 3-D and the following projection from 3-D to 2-D are called image warping.

A critical problem in DIBR is that regions occluded in the original view may become visible in the “virtual” view, an phenomenon also referred to as disocclusion. This is particularly problematic in case of extrapolation beyond the baseline of the original views as shown in Fig. 1. This figure shows an extrapolation scenario. One “virtual” camera view is extrapolated from one or two original ones. Scene areas that are visible in the original cameras are marked with blue lines. Those visible in the “virtual” view are highlighted with red lines. Disoccluded areas are visible in the “virtual” camera but invisible in the original ones and must therefore be reconstructed synthetically. The baseline is the distance between the projection centers of two cameras as shown in Fig. 1. A larger baseline between original and “virtual” camera leads to larger disocclusions.

In the absence of original image data, two extrapolation paradigms address this inherent disocclusion problem: 1) one can preprocess the depth information in a manner that no disocclusion occurs in the “virtual” view [26], or 2) replace the missing image areas (holes) with known suitable image information [27], [28], [29], [30]. In the following, a short overview will be given on relevant work in disocclusion handling in 3-D video. The presented methods were evaluated via subjective and objective tests. The results are presented in the following sections.

Fehn [26] pre-processes the per-pixel depth information with a 2-D Gaussian low-pass filter. The corresponding block diagram is depicted in Fig. 3. In this way large discontinuities

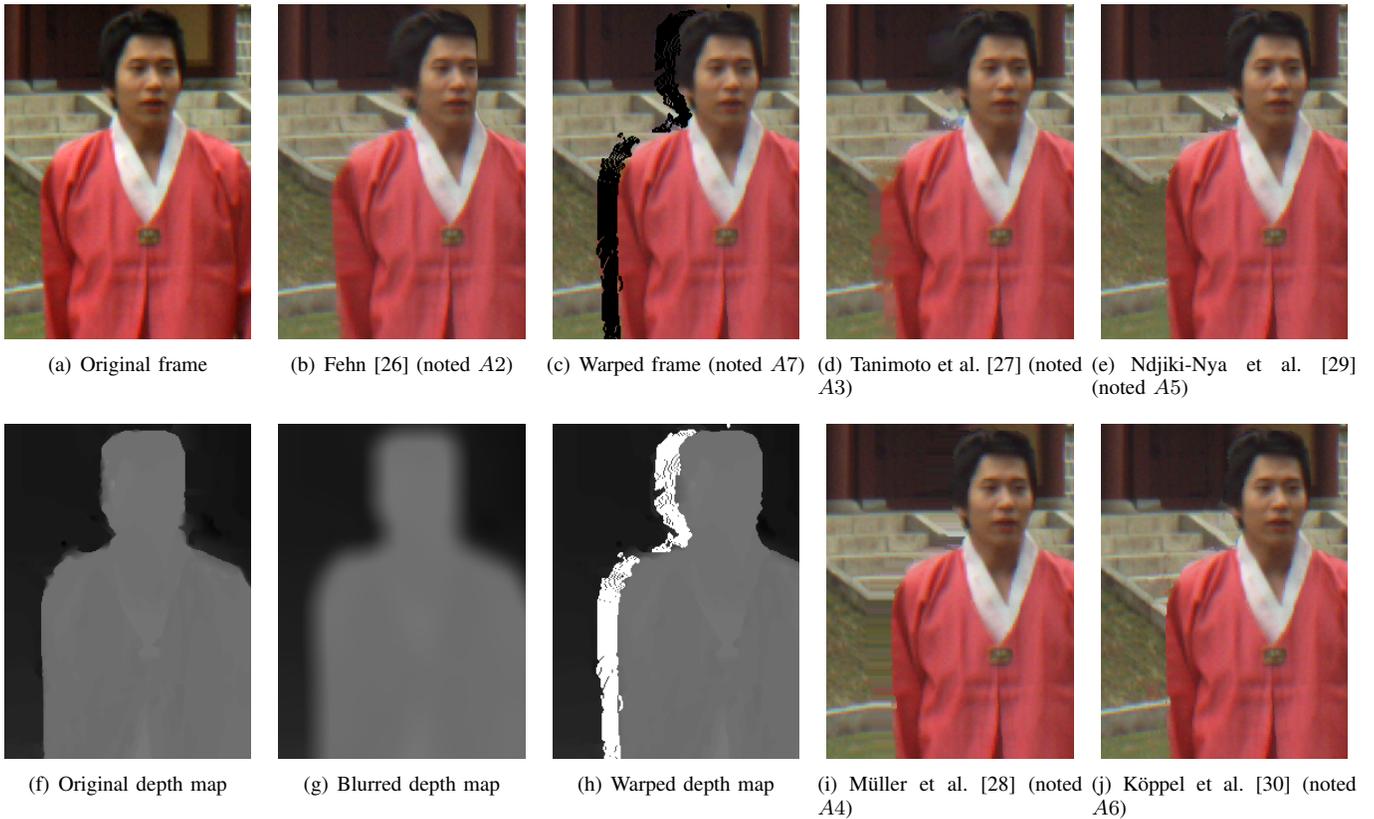


Fig. 2. DIBR results for frame 141 of the “Lovebird1” sequence. A baseline of 7cm is used. Camera 8 is used to render camera 6. (a) Original view. (f) Corresponding depth map. (b) “Virtual” camera view obtained with the method proposed by Fehn [26] using the Gaussian filtered depth map shown in (g). (c) Projected “virtual” view with disocclusions (marked black) using the original depth map (f). (h) Corresponding warped original depth map with disocclusions (marked white). (d) Result of the hole filling method proposed by Tanimoto et al. [27]. (e) Result of the hole filling method proposed by Ndjiki-Nya et al. [29]. (i) Result of the hole filling method proposed by Müller et al. [28]. (j) Result of the hole filling method proposed by Köppel et al. [30].

are smoothed out in the depth map and disocclusions do not appear in the “virtual” image. Fig. 2(b) shows a rendered result for a part of frame 141 of the sequence “Lovebird1”. The filtered depth map (cf. Fig. 2(g)) is used to warp the original view (cf. Fig. 2(a)) to the “virtual” position. The low-pass filtering of the depth map leads to geometric distortions in the “virtual” view (cf. Fig. 2(b)). The distortions are especially remarkable at background-foreground transitions. The distortions are visible on the stairway in the background which appears curved in the “virtual” view (cf. Fig. 2(b)). Also, the person in the foreground is widening to the left. In a rectified camera setup, the method proposed by Fehn [26] fails to extrapolate holes on the left or right image border. Therefore, these areas are dealt with in the following in two different ways. Either the border is cropped and the image is interpolated to its original size, or the holes on the border are inpainted using the method proposed by Telea [31]. These two methods are referred to as *A1* and *A2* respectively in the rest of the paper. The result shown in Fig. 2(b) is rendered by using method *A2*. The cropping method (*A1*) is only suitable for a stereo video, where only one view is transmitted and the other is rendered at the decoder side. In multi-view scenarios, *A1* is not applicable because all views (the original as well as the “virtual” views) have to be cropped to preserve the stereo impression, which would lead to image information losses in all the views. Furthermore, by cropping

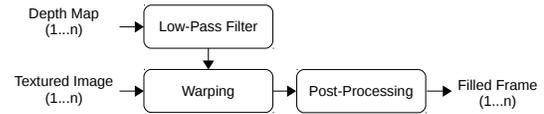


Fig. 3. Block diagram of the view synthesis approach proposed by Fehn [26], noted *A1*. The depth map is pre-processed with a 2-D Gaussian low-pass filter. In this way large discontinuities are smoothed out in the depth map. The filtered depth map is used to warp the original view to the “virtual” view. Holes in the left or right image border cannot be filled with this method in an extrapolation scenario. Therefore the holes at the borders are either inpainted [31] (algorithm *A2*), or the borders are cropped and the image is finally interpolated to its original size.

and subsequently interpolating the image, the distortions in the frame are enhanced.

Four methods that will be described hereafter, use the pixel-wise depth information (cf. Fig. 2(f)) to project the original frame (cf. Fig. 2(a)) to the “virtual” position. The projected image and the associated depth map with the disoccluded areas are shown in Fig. 2(c) and 2(h) respectively. The different methods fill in these holes in distinctive ways. Rendered results are shown for frame 141 of the “Lovebird1” sequence. Here, the original camera 8, is used to render camera 6.

Tanimoto et al. [27] proposed a 3-D view generation system that will be referred to as *A3* in the following sections. They use an inpainting method [31] to fill in missing parts

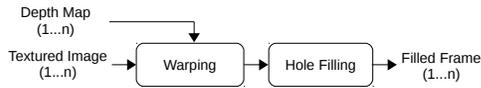


Fig. 4. Block diagram of the view synthesis approaches proposed by Tanimoto et al. [27] (noted *A3*) and Müller et al. [28] (noted *A4*) for a view extrapolation scenario. The pixel-wise depth information is used to warp the original image to the “virtual” view. Subsequently, Tanimoto et al. [27] cover the holes in the “virtual” image by using the inpainting method proposed in [31]. Müller et al. [28] fill the holes line-wise with neighboring background information.

in the “virtual” image. Their framework is primarily used to interpolate “virtual” views between two original cameras, where just small holes are to be closed. The block diagram for their extrapolation scenario is shown in Fig. 4. Fig. 2(d) shows a rendered result. The inpainting method [31] used to fill the holes (cf. Fig. 2(c)) introduces blur into the disoccluded area. In the extrapolation scenario the blurry regions are more obvious in the “virtual” view than in the interpolation case. Due to a missing foreground-background separation, the foreground image information is inserted into the disoccluded region. The inpainting method [31] fails to reconstruct complex textures. This algorithm was recently adopted as the reference software for MPEG standardization experiments in the 3-D Video group. First experiments have been conducted by MPEG to explore the extrapolation capabilities of the reference software [32].

Müller et al. [28] proposed a hole filling method embedded in a 3-D video system that will be referred to as *A4* in the following sections. The block diagram for the extrapolation scenario is shown in Fig. 4. Holes are filled line-wise with neighboring background information. The corresponding depth values at the hole boundary are examined row-wise to find background color samples to be copied into the hole. A rendered result is shown in Fig. 2(i). This method fails to reconstruct vertical or oblique structures and complex textures, as can be seen in the background of the rendered cut-out on the stairway. It is thus suitable for the reconstruction of simple and smooth non-textured backgrounds. Color extrapolation of the suitable background pixels leads to better results than a simple linear interpolation. Generally, due to depth estimation errors, some boundary background pixels may in fact belong to foreground objects. Thus, the color hints Müller et al. [28] exploit would lead to foreground color propagation into the hole.

In texture synthesis methods, the unknown regions are synthesized by copying content from the known parts of the image to the missing regions. Ndjiki-Nya et al. [29] proposed a hole filling approach for DIBR systems based on patch-based texture synthesis. This method is referred to as *A5* in the rest of the paper. The corresponding block diagram is shown in Fig. 5. Fig. 2(e) shows a rendered result. Holes with small spatial extent are closed by solving Laplacian equations. Larger holes are initialized by median filtering and then optimized via texture synthesis. Thanks to the usage of texture synthesis, it is possible to reconstruct complex textures, as can be seen in the background area in Fig. 2(e). The used patches are rectangular, which may lead to block artifacts. As the known

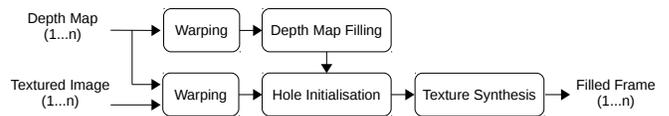


Fig. 5. Block diagram of the view synthesis approach proposed by Ndjiki-Nya et al. [29] (noted *A5*). The original image and the associated pixel-wise depth information is warped to the original view. Then the disocclusions in the depth map are filled. Finally, the holes in the image are initialized and subsequently refined with texture synthesis.

background area is finite and may not be a good representative of the texture to be predicted, even the best match may be a bad compromise and lead to artifacts in the synthesized texture. Due to the utilized texture synthesis approach [29] only straight edges can be accurately reconstructed.

The presented view synthesis methods *A1-A5* render a sequence frame-by-frame. Information from previous or subsequent frames is not considered. This can lead to temporal inconsistencies such as flickering.

Köppel et al. [30] extended the *A5* approach by a background sprite. The sprite stores valuable background image information and is updated frame-wise. Using the original and synthesized image information from previous frames, temporal consistency is achieved in a sequence. A rendered result is shown in Fig. 2(j) and the block diagram is depicted in Fig. 6. Background texture is copied from the sprite to the disoccluded area so that background details are preserved. The image information from previous frames is selected via the pixel-wise depth information. However, the latter not always reliable. Therefore, incorrectly copied background image information can degrade the quality of the “virtual” view. Also, the blocking artifacts described for method *A5* can occur as well. In subjective tests conducted in section III only images are evaluated. Thus, the capabilities of the approach [30] to achieve temporal consistency in a sequence are not investigated in this work. This method is referred to as *A6* in the rest of the paper.

Unfilled sequences (i.e. those with holes) are referred to as *A7* in the rest of the paper (cf. Fig. 2(c)).

### III. SUBJECTIVE MEASUREMENTS

The objective of the experiments is to determine whether two efficient protocols of subjective assessment of quality (in a 2-D context) are still adapted to assess the quality of DIBR-based synthesized images. Three test sequences have been used to generate four different viewpoints, that is to say twelve synthesized sequences for each tested algorithm (84 synthesized sequences in total): *Book Arrival* (1024×768, 16 cameras with 6.5-cm spacing), *Lovebird1* (1024×768, 12 cameras with 3.5-cm spacing) and *Newspaper* (1024×768, 9 cameras with 5-cm spacing). Altogether, 43 naive (i.e., non-expert) observers participated in the subjective assessment session. They were mostly computer science students with no prior knowledge of the sequences. The session was conducted in an ITU conforming test environment. Absolute categorical rating (ACR) [33] and Paired comparisons [34] were used to collect perceived quality scores. The stimuli were displayed

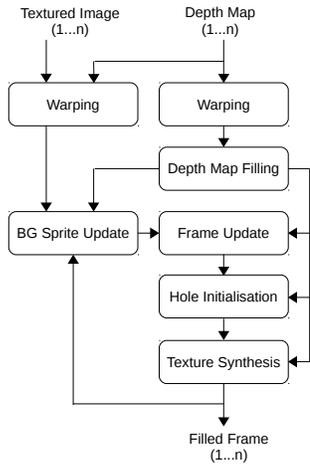


Fig. 6. Block diagram of the view synthesis approach proposed by Köppel et al. [30] (noted A6). The original view and the associated depth map are projected to the “virtual” position. Next, the background sprite is updated with original background data and the holes in the “virtual” image are updated with background information from the sprite. The remaining holes are first initialized and subsequently refined with texture synthesis.

on a TVLogic LVM401W, and according to ITU-T BT.500 [35]. Considering the large size of the tested database, only key frames were presented to the observers, since still images can also be a plausible scenario for FVV.

#### A. Subjective Methodologies

As explained in the introduction, subjective quality assessment is delicate when addressing new type of conditions because one has to define the optimal way to get reliable data. New conditions are introduced, and the right protocol to assess the visual quality with observers is still an open question. Recently, authors considered these new conditions and concepts, such as the presence, or the perception of depth, in [36] and [2]. In the absence of any better subjective quality assessment methodologies for the evaluation of synthesized views, the measurement is mostly obtained through 2-D validated protocols.

*Absolute Categorical Rating Method* : The absolute categorical rating (ACR) [37] method consists in presenting several test conditions (also called trials) to observers. A single trial is presented only once and the all the trials are presented in a random order. Observers score the test item according to a discrete category rating scale. The five-level (ACR-5) used scale is shown in Table I. Each adjective corresponds to an integer ranging from 1 to 5. Observers’ opinion scores are then averaged by computing a mean opinion score (MOS). ACR requires many observers to minimize the contextual effects (previously presented stimuli influence the observers’ opinion, i.e. presentation order influences opinion ratings).

*Paired Comparisons (PC) Method*: With this method, a pair of images is presented to the observer who selects the one that best satisfies the specified judgment criterion, i.e. the best image quality. Responses from many observers yield to an interval-scale ordering of images: they can be converted to scale values using Thurstone-Mosteller’s or Bradley-Terry’s

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

TABLE I  
COMPARISON SCALE FOR ACR

model [38]. This leads to a hypothetical perceptual continuum. The presented experiments follow Thurstone-Mosteller’s model where naive observers were asked to choose the preferred item from one pair. Paired comparisons are time consuming: for  $N$  items,  $\frac{N(N-1)}{2}$  comparisons are executed.

#### B. Suitability of Usual Subjective Assessment Protocols

Statistical analyses have been conducted over subjective measurements. A Student’s t-test has been performed over the MOS scores, and over the PC scores for each algorithm. This is meant to provide knowledge on the statistical equivalence of the algorithms. Tables II and III show the results of the statistical tests over MOS and PC values respectively. In both tables, the number in parentheses indicates the required minimum number of observers that allow the statistical distinction (VQEG recommends 24 participants as a minimum, values in bold are higher than 24 in the tables). Table III indicates that only A2-A4 and A2-A6 are statistically equivalent. Besides, this is confirmed by Table II, because the same groups are identified among others (A5-A2, A6-A4 for example). In the case of PC test, most of the algorithms can be statistically distinguished with less than 24 participants. However, in the case of ACR results, the final distinction seems stable when 32 observers participate. This suggests that fewer observers are needed for a PC-based subjective assessment. However, this method is very time-consuming: first because of the large number of comparisons; and second, more observers are required in order to avoid the contextual effect. For these experiments, in average, one observer needed 45 minutes to achieve the PC test but only 15 minutes for the ACR test.

## IV. OBJECTIVE MEASUREMENTS

The objective of the experiments is to assess the reliability of objective metrics in the case of images synthesized from DIBR algorithms, by determining the tested algorithms performances. The same test material as in subjective measurements (Section III) was used in these objective measurements. Key frames were evaluated through different objective metrics through MeTriX MuX Visual Quality Assessment Package [39]. For both objective metrics, the reference was the original acquired image.

#### A. Objective Tools

Different objective metrics are used in the presented experiments:

- *Peak Signal-to-Noise Ratio (PSNR)* measures the signal fidelity of a distorted image compared to a reference. It is based on the measure of the mean squared error (MSE).

- As an alternative to pixel-based methods, universal quality index *UQI* [40] is a perceptual-like metric. It models the image distortion by a combination of three factors: loss of correlation, luminance distortion, and contrast distortion.
- *PSNR-HVS* [41], based on PSNR and UQI, is meant to take into account the Human Visual System (HVS) properties. *PSNR-HVSM* [42] is based on PSNR but takes into account contrast sensitivity function (CSF) and between-coefficient contrast masking of DCT basis functions.
- *Single-scale Structural SIMilarity (SSIM)* [43] combines image structural information: mean, variance, covariance of pixels, for a single local patch. *Multi-scale SSIM (MSSIM)* is the average SSIM scores of all the patches of the image.
- *Visual Signal-to-Noise Ratio (VSNR)* [44] is also a perceptual-like metric: it is based on a visual detection of distortion criterion, helped by CSF.
- *Weighted Signal-to-Noise Ratio (WSNR)* that uses a weighting function adapted to HVS, denotes a weighted signal-to-noise ratio.
- *Visual Information Fidelity (VIF)* [45] uses a statistical model and an image distortion model to assess an image. *VIFP* is a pixel-based version of VIF.
- *Information Fidelity Criterion (IFC)* [46] uses a distortion model to evaluate the information shared between the reference image and the degraded image.
- *Noise quality measure (NQM)* [47] quantifies the injected noise in the tested image.

### B. Correlation Between Objective and Subjective Measures

From the whole set of tested data, measures yield the performances of the algorithms. Objective and subjective measurements lead to different rankings, depending on the metric, as showed on Table IV. Subjective-based rankings point A1 as the best out of the seven algorithms. But objective-based rankings judge it as the worst. Similarly, A6 obtains the best results for numerous objective metrics, but subjective assessments judge the synthesized views as “poor”. A2, A3, A4 and A5 ranks seem to be relatively coherent with subjective evaluations. This table suggests that objective metrics detect non-annoying artifacts. Another remark concerns the fact that the objective metrics rankings are very close to each other. This suggests that the tested objective metrics are correlated. This is confirmed in Table V. This table also shows the high correlation with pixel-based and more perceptual-like metrics such as PSNR and SSIM (83.9%).

Results also show that the algorithm performances are linked to the sequences properties: Fig. 7 shows the MOS scores from ACR and Fig. 8 shows the scaled values from PC, obtained for each algorithm, depending on the sequence. Algorithms are ranked by obtained scores. The ranks vary depending on the tested sequence. A4 seems to perform better on *Lovebird1* than the other algorithms, and it is ranked as the best algorithm but only for this sequence. However, note that all the scores for this sequence are lower than those of the

other sequences, regardless the algorithm. This suggests that the algorithms performances are tied to the sequence content or corroborates [9] concerning the influence of the content on human opinion. The same trend is observed with PC results, and with the objective measures.

The consistency between objective and subjective measures is evaluated by computing the correlation coefficients for the all the fitted measured points. Table VI show the results. Note that none of the tested metric is 50% close to human judgment when assessing synthesized views. According to MOS and PC scores, the most correlated metrics are  $PSNR_{HVSM}$  (38.1% and 42.2% resp.),  $PSNR_{HSV}$  (37.3% and 41.9% resp.),  $PSNR$  (38.6% and 40.0% resp.) and  $WSNR$  (42.3% and 36.9% resp.). These are pixel-based or perceptual-like metrics. In order to further consider the consistency, Fig. 9 shows the ACR values over the  $PSNR_{HVSM}$  scores. The figure shows that a difference of nearly 9 dB (A1 and A5) leads to nearly the same absolute rate.

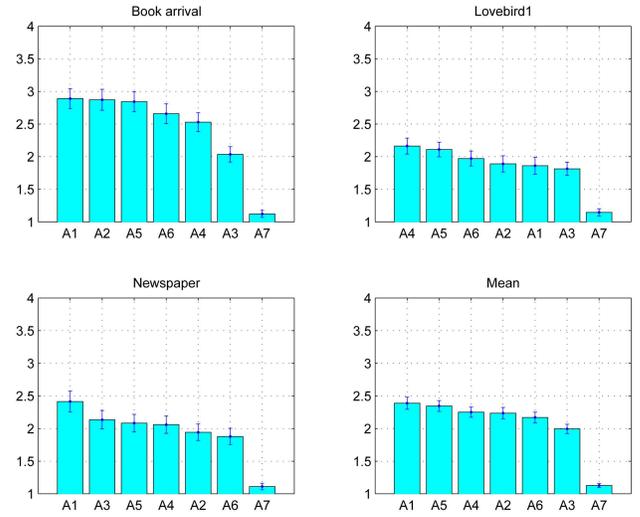


Fig. 7. Mean ACR Scores for the different sequences.

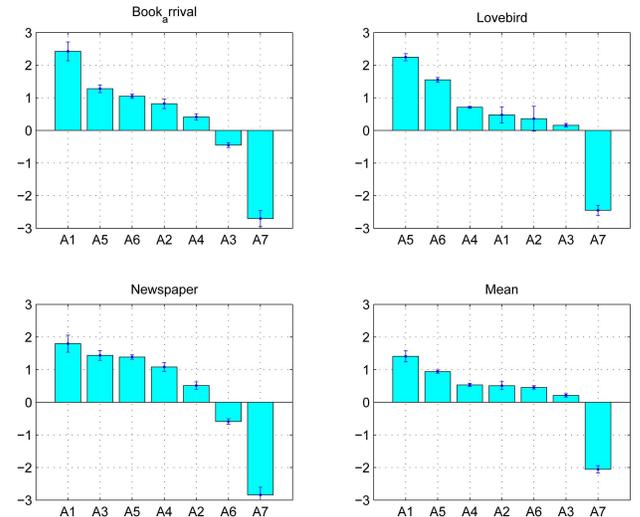


Fig. 8. Paired Comparisons scaled values for the different sequences.

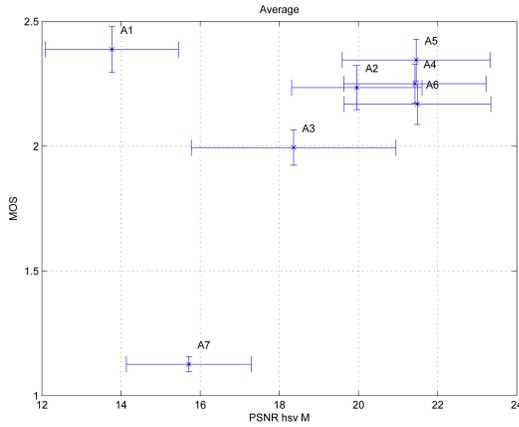


Fig. 9. Correlation between ACR and  $PSNR_{HSM}$ .

## V. FUTURE DIRECTIONS

These results point out the lack of crucial information that would help the assessment of synthesized views. In this section, we propose an analysis of the location of the artifacts and the first results for new objective assessment tools.

### A. Analysis of Artifacts

Comparing the algorithms performances depending on the baseline distance, it can be observed that the shorter the baseline, the better the scores are, as expected (for both subjective and objective scores). Also, the algorithm rankings remain stable from short to larger baseline distance (large distance is defined as twice the acquisition camera baseline distance). This suggests that each algorithm induces a specific type of distortion that is located in the disoccluded areas while the rest of the image remains close or relatively perceptually acceptable to humans.

The analysis of the location of the artifacts, illustrated in Fig. 10 confirms the assumption concerning the close relationship between algorithms and distortion types. The images are obtained by computing the difference between original and synthesized view, and then applying a threshold  $Th$  to detect the most important errors (identified as critical areas). It is defined as follows, to detect only the most important errors in the synthesized image:

$$Th = \frac{\max(I-I')}{10}$$

where  $I$  is the original image, and  $I'$  is the synthesized view.

In our case, unlike the commonly encountered artifacts in video coding, artifacts are mainly located around the disoccluded areas, while the rest of the image remains close or the same as the original. Most of the video coding standards rely on a block-based DCT, and the resulting artifacts are often scattered in the whole image, as studied in [48]. Artifacts resulting from DIBR algorithms are different and located in specific areas.

As explained in Section II, the critical problem in DIBR is the disocclusions that are more likely to occur around sharp discontinuities. Thus the artifacts are located in these specific areas. By extrapolating the missing areas, the tested algorithms

induce errors in these areas.  $A1$ ,  $A2$  and  $A3$  induce blur in these regions, making the background-foreground transitions unnatural. The distortions seem to be even more visible and annoying if the color contrast between foreground and background is important, which make these cases critical. The best methods, in the sense of objective metrics are  $A4$ ,  $A5$  and  $A6$ . All three show fewer errors around the disoccluded areas (Fig. 10). However their most significant errors are located in the text areas. All the objective metrics penalize  $A1$  and  $A3$  because the count numerous errors around the disoccluded areas.

So, the type and the location of artifacts are specific in the case of DIBR-based synthesized images, and are different from the current cases of 2-D images. Here, a region may be slightly transferred or resized, depending on the chosen extrapolation method (if the method chooses to assign the background values to the missing areas, object may be resized). If the region is still visually coherent, subjective scores should be correct, but pixel-based metrics should penalize this type of image (case of  $A1$ ). For this reason, we believe that a novel assessment metric should consider the degradations of the image coupled with a penalizing factor associated to the transformation of the critical areas. This, in order to measure the amount of geometric distortions.

### B. Analysis of the Contours Consistency

Our assumption concerning the specific location of artifacts around the disoccluded areas, and more precisely around strong depth discontinuities, led us to consider the contours of the synthesized view. Then, we propose to penalize important shifts of contours. The block diagram in Fig. 11 illustrates the principles of this approach.

As illustrated in Fig. 11, we first apply a Canny Edge Detector on the reference image and on the synthesized view in order to detect their contours. Then a motion vectors estimation is computed over the contours of the synthesized view, in comparison with the contours of the reference view. Considering a neighborhood of pixel consisting of a  $3 \times 3$  window, neighbor vectors whose angle with the motion vector of the central pixel is up to  $45^\circ$  are considered as deviant and central pixel is penalized by a vote from each deviant vector. Then, the score of the central pixel is the sum of the votes of each neighbor vectors. The final quality score of the image is obtained by averaging all the votes computed previously for each pixel.

As pointed out in Section IV, the thin consistency between objective and subjective measures was illustrated by two distinct items obtaining the same subjective score but two very different objective scores; or two distinct items obtaining the same objective score but two very different subjective scores. We considered these four items. As a preliminary experiment, we then assessed the quality of four critical items with an objective metric (SSIM) and with the proposed measure. Results are plotted in Fig. 12. Concerned about the clarity of the presented results, we scaled the points obtained with the proposed approach, according to the same affine transform: we first scaled two of the obtained points

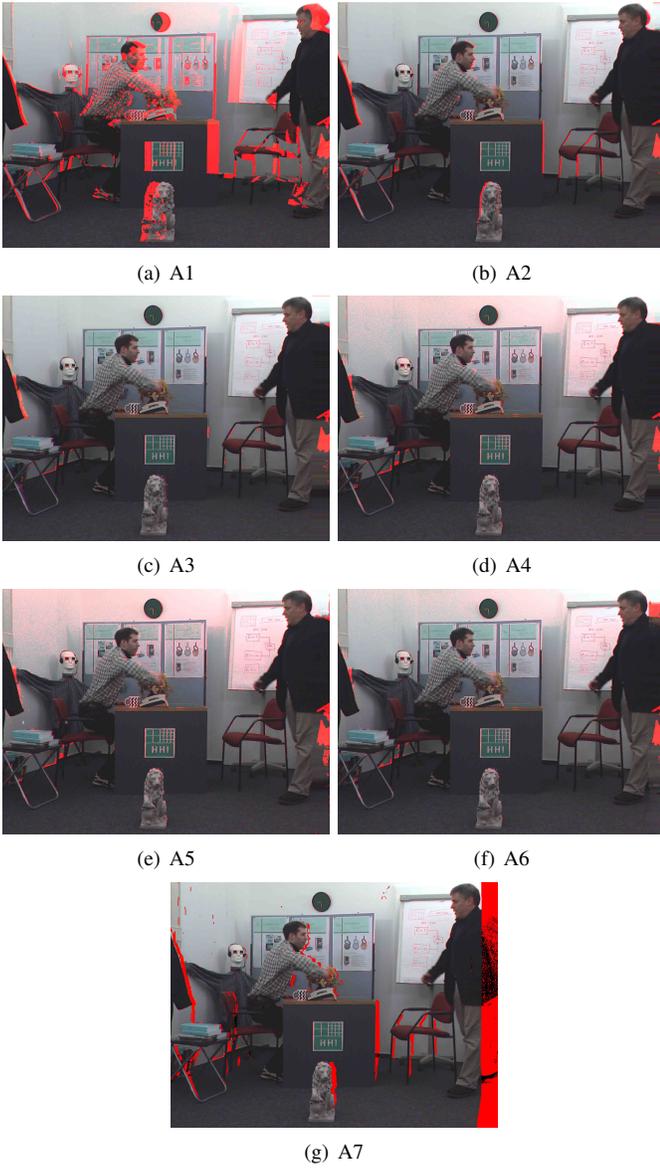


Fig. 10. Location of most important errors depending on used synthesis algorithm (in red).

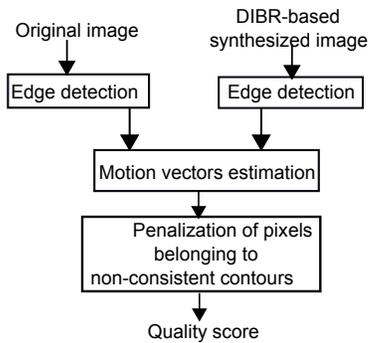


Fig. 11. Block diagram of the proposed approach in Section V-B.

so that they lay on the corresponding scores of the SSIM; then we applied the same obtained transformation to the two other points. The arrows on the figure, illustrate the fact that the proposed approach moves the points closer to a linear

regression line. The consistency with the subjective scores is improved. The correlation coefficients are 18% and 84% for SSIM and proposed approach respectively.

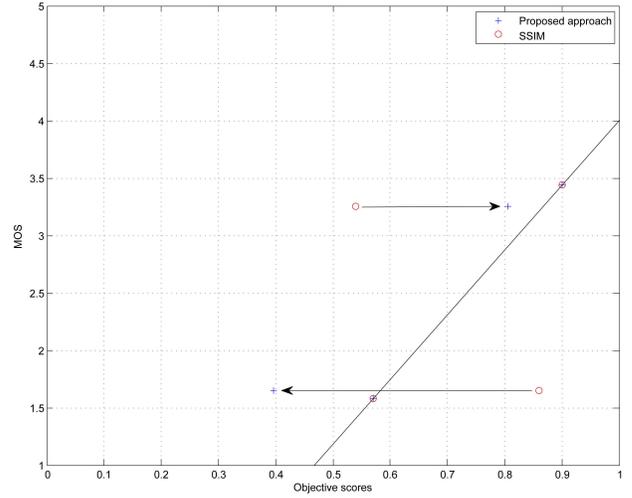


Fig. 12. MOS scores over scores of objective metrics (SSIM and proposed approach in Section V-B).

Considering the consistency of the contour shifts, these very first results are encouraging and are to be investigated in future work.

### C. Analysis of the distorted areas

In a second approach, based on the observations mentioned in Section V-A, we considered the distorted areas. Fig. 13 shows the block diagram of the proposed approach. It consists in applying an objective metric only on the critical areas. The critical areas are defined by the main differences between the reference view and the synthesized view, as mentioned in Section V-A. Then we apply SSIM measure on these areas and the final score is the mean SSIM scores normalized by the amount of distorted pixels.

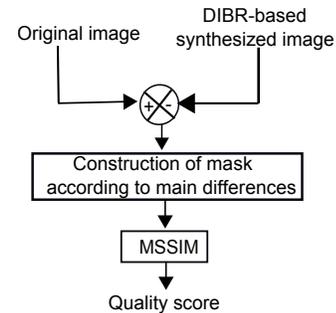


Fig. 13. Block diagram of the proposed approach in Section V-C.

As previously, we assessed the quality of four critical items with an objective metric (SSIM) and with the proposed measure. Results are plotted in Fig. 14. As explained above, we scaled the points obtained with the proposed approach. The figure shows that the consistency with subjective scores is slightly improved with the proposed approach. The correlation

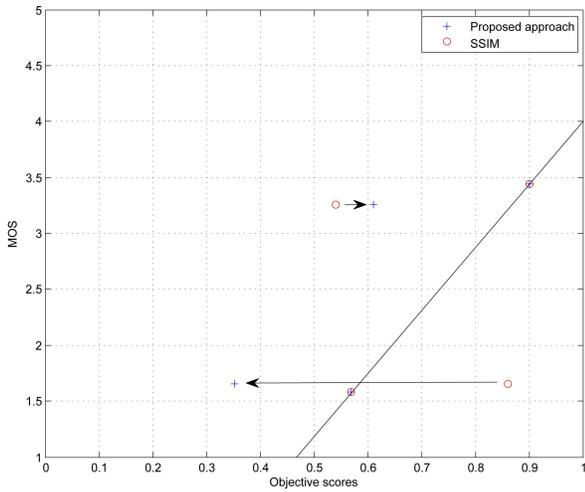


Fig. 14. MOS scores over scores of objective metrics (SSIM and proposed approach in Section V-C).

coefficients are 18% and 78% for SSIM and the proposed approach respectively.

The observation of the disoccluded areas into an objective measure seems encouraging but further investigation should be run.

## VI. CONCLUSION

Subjective measurements (ACR and Paired comparisons) were coupled with objective measurements in the presented experiments, assessing views synthesized from seven different algorithms. The experiments considered the commonly used methods for assessing conventional images, subjectively as well as objectively. Concerning the subjective evaluations, Paired comparisons and ACR results highly correlate. But statistical analyses show that fewer observers were required for Paired comparisons tests to establish the algorithms distinctions. However, this is a time-consuming method, often avoided by researchers. Moreover, when the number of items to be compared is high, the test is hardly feasible. Concerning the objective metrics, the results show that usual objective assessments hardly correlate with subjective assessments. Rankings of algorithms from objective metrics are not reliable, when regarding the subjective results. The presented experiments revealed that using only the objective metrics is not sufficient for assessing virtual synthesized views, though they give information on the presence of errors. The analysis of their capabilities coupled with the analysis of the distortions induced by view synthesis might help in designing a new metric. We presented the first results of two measurement approaches. The first proposal was based on the analysis of the orientation of the edge shifts. The second proposal was based on the analysis of the disoccluded areas. Both approaches gave encouraging results when considering items deviant with usual 2D metrics. The correlation with the subjective measurements is encouraging. Further experiments should be performed on video sequences in the future, to refine the presented results.

	A1	A2	A3	A4	A5	A6	A7
A1		↑(32)	↑(<24)	↑(32)	o (>43)	↑(30)	↑(<24)
A2	↓(32)		↑(<24)	o (>43)	o (>43)	o (>43)	↑(<24)
A3	↓(<24)	↓(<24)		↓(<24)	↓(<24)	↓(<24)	↑(<24)
A4	↓(32)	o (>43)	↑(<24)		o (>43)	o (>43)	↑(<24)
A5	o (>43)	o (>43)	↑(<24)	o (>43)		↑(28)	↑(<24)
A6	↓(30)	o (>43)	↑(<24)	o (>43)	↓(28)		↑(<24)
A7	↓(<24)	↓(<24)	↓(<24)	↓(<24)	↓(<24)	↓(<24)	

TABLE II  
RESULTS OF STUDENT'S T-TEST WITH ACR RESULTS. LEGEND:↑: SUPERIOR, ↓: INFERIOR, o: STATISTICALLY EQUIVALENT. READING: LINE"1" IS STATISTICALLY SUPERIOR TO COLUMN "2". DISTINCTION IS STABLE WHEN "32" OBSERVERS PARTICIPATE.

	A1	A2	A3	A4	A5	A6	A7
A1		↑(<24)	↑(<24)	↑(<24)	↑(<24)	↑(<24)	↑(<24)
A2	↓(<24)		↑(28)	o (<24)	↓(<24)	o (>43)	↑(<24)
A3	↓(<24)	↓(28)		↓(<24)	↓(<24)	↓(<24)	↑(<24)
A4	↓(<24)	o (>43)	↑(<24)		↓(<24)	↑(43)	↑(<24)
A5	↓(<24)	↑(<24)	↑(<24)	↑(<24)		↑(<24)	↑(<24)
A6	↓(<24)	o (>43)	↑(<24)	↓(<43)	↓(<24)		↑(<24)
A7	↓(<24)	↓(<24)	↓(<24)	↓(<24)	↓(<24)	↓(<24)	

TABLE III  
RESULTS OF STUDENT'S T-TEST WITH PAIRED COMPARISONS RESULTS. LEGEND:↑: SUPERIOR, ↓: INFERIOR, o: STATISTICALLY EQUIVALENT. READING: LINE"1" IS STATISTICALLY SUPERIOR TO COLUMN "2". DISTINCTION IS STABLE WHEN "LESS THAN 24" OBSERVERS PARTICIPATE.

	A1	A2	A3	A4	A5	A6	A7
MOS	2.388	2.234	1.994	2.250	2.345	2.169	1.126
Rank order	1	4	6	3	2	5	7
PC	1.4038	0.5081	0.2073	0.5311	0.9363	0.4540	-2.0547
Rank order	1	4	6	3	2	5	7
PSNR	18.752	24.998	23.180	26.117	26.171	26.177	20.307
Rank order	7	4	5	3	2	1	6
SSIM	0.638	0.843	0.786	0.859	0.859	0.858	0.821
Rank order	7	4	6	1	1	3	5
MSSIM	0.648	0.932	0.826	0.950	0.949	0.949	0.883
Rank order	7	4	6	1	2	2	5
VSNR	13.135	20.530	18.901	22.004	22.247	22.195	21.055
Rank order	7	5	6	3	1	2	4
VIF	0.124	0.394	0.314	0.425	0.425	0.426	0.397
Rank order	7	5	6	2	2	1	4
VIFP	0.147	0.416	0.344	0.448	0.448	0.448	0.420
Rank order	7	5	6	1	1	1	4
UQI	0.237	0.556	0.474	0.577	0.576	0.577	0.558
Rank order	7	5	6	1	3	1	4
IFC	0.757	2.420	1.959	2.587	2.586	2.591	2.423
Rank order	7	5	6	2	3	1	4
NQM	8.713	16.334	13.645	17.074	17.198	17.201	10.291
Rank order	7	4	5	3	2	1	6
WSNR	13.817	20.593	18.517	21.597	21.697	21.716	15.588
Rank order	7	4	5	3	2	1	6
PSNR HSV	13.772	19.959	18.362	21.428	21.458	21.491	15.714
Rank order	7	4	5	3	2	1	6
PSNR HSV	13.530	19.512	17.953	20.938	20.958	20.987	15.407
Rank order	7	4	5	3	2	1	6

TABLE IV  
RANKINGS ACCORDING TO MEASUREMENTS

## ACKNOWLEDGMENT

This work is partly supported by French National Research Agency as part of ANR-PERSEE project (ANR-09-BLAN-0170), and ANR-CAIMAN (ANR-08-VERS-002) project.

## REFERENCES

- [1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3-D video and free viewpoint Video-Technologies, appli-

	PSNR	SSIM	MSSIM	VSNR	VIF	VIFP	UQI	IFC	NQM	WSNR	PSNR <sub>hsvm</sub>	PSNR <sub>hsv</sub>
PSNR		83.9	79.6	87.3	77.0	70.6	53.6	71.6	95.2	98.2	99.2	99.0
SSIM	83.9		96.7	93.9	93.4	92.4	81.5	92.9	84.9	83.7	83.2	83.5
MSSIM	79.6	96.7		89.7	88.8	90.2	86.3	89.4	85.6	81.1	77.9	78.3
VSNR	87.3	93.9	89.7		87.9	83.3	71.9	84.0	85.3	85.5	86.1	85.8
VIF	77.0	93.4	88.8	87.9		97.5	75.2	98.7	74.4	78.1	79.4	80.2
VIFP	70.6	92.4	90.2	83.3	97.5		85.9	99.2	73.6	75.0	72.2	72.9
UQI	53.6	81.5	86.3	71.9	75.2	85.9		81.9	70.2	61.8	50.9	50.8
IFC	71.6	92.9	89.4	84.0	98.7	99.2	81.9		72.8	74.4	73.5	74.4
NQM	95.2	84.9	85.6	85.3	74.4	73.6	70.2	72.8		97.1	92.3	91.8
WSNR	98.2	83.7	81.1	85.5	78.1	75.0	61.8	74.4	97.1		97.4	97.1
PSNR <sub>hsvm</sub>	99.2	83.2	77.9	86.1	79.4	72.2	50.9	73.5	92.3	97.4		99.9
PSNR <sub>hsv</sub>	99.0	83.5	78.3	85.8	80.2	72.9	50.8	74.4	91.8	97.1	99.9	

TABLE V  
CORRELATION COEFFICIENTS BETWEEN OBJECTIVE METRICS IN PERCENTAGE.

	PSNR	SSIM	MSSIM	VSNR	VIF	VIFP	UQI	IFC	NQM	WSNR	PSNR <sub>hsvm</sub>	PSNR <sub>hsv</sub>
CCMOS	38.6	21.9	16.1	25.8	19.3	19.2	20.2	19.0	38.6	42.3	38.1	37.3
CCPC	40.0	23.8	34.9	19.7	16.2	22.0	32.9	20.1	37.8	36.9	42.2	41.9

TABLE VI  
CORRELATION COEFFICIENTS BETWEEN MOS AND OBJECTIVE SCORES IN PERCENTAGE.

cations and MPEG standards,” in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME06)*, 2006, pp. 2161–2164.

[2] P. Seuntjens, “Visual experience of 3D TV,” *doctor doctoral thesis, Eindhoven University of Technology*, 2006.

[3] M. S Banks, K. Akeley, D. M Hoffman, and A. R Girshick, “Consequences of incorrect focus cues in stereo displays,” *Journal of the Society for Information Display*, vol. 24, no. 7, pp. 7, 2008.

[4] W. J Tam, “Human stereoscopic vision: Research applications for 3D-TV,” vol. 1, no. 38, pp. 1216–1219, 2007.

[5] M. Meesters, W. Ijsselstein, and P. Seuntjens, “A survey of perceptual evaluations and requirements of three dimensional TV,” *IEEE Transactions on Circuits And Systems for Video Technology*, vol. 14, no. 3, pp. 381–391, Mar. 2004.

[6] A. Boev, A. Gotchev, and K. Egiuzarian, “Stereoscopic artifacts on portable auto-stereoscopic displays: what matters,” *Proc. of VPQM*, 2009.

[7] M. Cancellaro, V. Palma, and A. Neri, “Stereo video artifacts introduced by a distributed coding approach,” in *Fifth International Workshop on Video Processing and Quality Metrics (VPQM)*, Arizona, U.S.A., 2010.

[8] Noha A. El-Yamany, Kemal Ugur, Miska M. Hannuksela, and Moncef Gabbouj, “Evaluation of depth compression and view synthesis distortions in multiview-video-plus-depth coding systems,” in *2010 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, Tampere, Finland, 2010, pp. 1–4.

[9] Q. Huynh-Thu, P. Le Callet, and M. Barkowsky, “Video quality assessment: From 2-D to 3-D-Challenges and future trends,” in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, pp. 4025–4028.

[10] Y. Pan, I. Cheng, and A. Basu, “Quality metric for approximating subjective evaluation of 3-D objects,” *Multimedia, IEEE Transactions on*, vol. 7, no. 2, pp. 269279, 2005.

[11] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, “New requirements of subjective video quality assessment methodologies for 3DTV,” in *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2010*, Scottsdale, Arizona, U.S.A., 2010.

[12] I. Cheng and P. Boulanger, “A 3d perceptual metric using just-noticeable-difference,” in *Proceedings of the Annual Conference of the European Association for Computer Graphics (Eurographics’ 05)*, p. 97100.

[13] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B Akar, G. Triantafyllidis, and A. Koz, “Coding algorithms for 3DTV-a survey,” *IEEE transactions on circuits and systems for video technology*, vol. 17, no. 11, pp. 1606–1620, 2007.

[14] A. Boev, A. Gotchev, K. Egiuzarian, A. Aksay, and G. B Akar, “Towards compound stereo-video quality metric: a specific encoder-based frame-

work,” in *Proc. Southwest Symp. Image Analysis and Interpretation (SSIAI 2006)*, 2006, pp. 218–222.

[15] R. Patterson, “Human factors of 3-D displays,” *Journal of the SID*, vol. 15, no. 11, pp. 861–871, 2007.

[16] S. T Na, K. J Oh, and Y. S Ho, “Joint coding of multi-view video and corresponding depth map,” in *Proc. IEEE International Conference on Image Processing (ICIP’08)*, San Diego, CA, USA, 2008.

[17] Y. Morvan, D. Farin, and P.H.N. de With, “Joint depth/texture bit-allocation for multi-view video compression,” in *Proceedings of Picture Coding Symposium (PCS 2007)*, Lisboa, Portugal, Nov. 2007, vol. 10, p. 4349.

[18] ISO/IEC JTC1/SC29/WG11 MPEG doc. no. N11831, “Description of exploration experiments in 3-D video coding,” Daegu, Korea, Jan. 2011.

[19] “VQEG 3DTV Group,” <http://www.its.bldrdoc.gov/vqeg/projects/3dtv/>.

[20] J. You, L. Xing, A. Perkis, and X. Wang, “Perceptual quality assessment for stereoscopic images based on 2-D image quality metrics and disparity analysis,” in *Proc. Int. Workshop Video Processing and Quality Metrics, Scottsdale, Arizona, USA*, 2010.

[21] S. L. P. Yasakethu, C. Hewage, W. Fernando, and A. Kondoz, “Quality analysis for 3-D video using 2-D video quality models,” *Consumer Electronics, IEEE Transactions on*, vol. 54, no. 4, pp. 1969–1976, 2008.

[22] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Müller, “Quality assessment of 3D video in rate allocation experiments,” in *IEEE Int. Symposium on Consumer Electronics (14-16 April, Algarve, Portugal)*, 2008.

[23] K. T. Mullen, “The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings,” *The Journal of Physiology*, vol. 359, no. 1, pp. 381, 1985.

[24] S. A Klein, T. Carney, L. Barghout-Stein, and C. W Tyler, “Seven models of masking,” in *Proceedings of SPIE*, 1997, vol. 3016, p. 13.

[25] W. F Schreiber, *Fundamentals of electronic imaging systems: some aspects of image processing*, Springer-Verlag, 1993.

[26] C. Fehn, “Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV,” in *Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems X*, San Jose, USA, January 2004.

[27] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, “View generation with 3-D warping using depth information for FTV,” *Elsevier Signal Processing: Image Communication*, vol. 24, pp. 65–72, 2009.

[28] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, “View synthesis for advanced 3-D video systems,” *EURASIP Journal on Image and Video Processing*, 2008, Article ID 438148, 11 pages.

[29] P. Ndjiki-Nya, M. Köppel, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand, “Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering,” in *Proc. IEEE International Conference on Multimedia & Expo (ICME)*, Singapore, July 2010.

[30] M. Köppel, P. Ndjiki-Nya, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand, “Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering,” in *Proc. IEEE International Conference on Image Processing (ICIP)*, Hong Kong, China, September 2010.

[31] A. Telea, “An image inpainting technique based on the fast marching method,” *Journal of Graphics, GPU, and Game Tools*, vol. 9, no. 1, pp. 23–34, 2004.

[32] Output document, “Description of exploration experiments in 3-D video coding,” ISO/IEC/SC29/WG11 MPEG2010/N11630, October 2010, Guangzhou, China.

[33] ITU-T Study Group 12, “ITU-T p.910 subjective video quality assessment methods for multimedia applications,” 1997.

[34] B. W Keelan, *Handbook of image quality: characterization and prediction*, CRC, 2002.

[35] ITU-R BT., 500, *Methodology for the subjective assessment of the quality of television pictures*, November, 1993.

[36] W. Ijsselstein, H. Ridder, J. Freeman, S. E. Avons, and D. Bouwhuis, “Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence,” *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 298311, 2001.

[37] International Telecommunication Union (ITU) Radiocommunication Assembly, “Subjective video quality assessment methods for multimedia applications,” 2008.

[38] J. C Handley, “Comparative analysis of Bradley-Terry and Thurstone-Mosteller paired comparison models for image quality assessment,” in *ISand TS PICS Conference*, 2001, pp. 108–112.

[39] “MetriX MuXpage,” [http://foulard.ece.cornell.edu/gaubatz/metrix\\_mux/](http://foulard.ece.cornell.edu/gaubatz/metrix_mux/).

[40] Z. Wang and A. C Bovik, “A universal image quality index,” *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81–84, 2002.

- [41] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS," in *CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, USA, 2006.
- [42] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *CD-ROM Proc. of the Third International Workshop on Video Processing and Quality Metrics*, 2007, vol. 4.
- [43] Z. Wang, A. C Bovik, H. R Sheikh, and E. P Simoncelli, "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.
- [44] D. M Chandler and S. S Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images," *Image Processing, IEEE Transactions on*, vol. 16, no. 9, pp. 2284–2298, 2007.
- [45] H. R Sheikh and A. C Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, 2006.
- [46] H. R Sheikh, A. C Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2117–2128, 2005.
- [47] N. Damera-Venkata, T. D Kite, W. S Geisler, B. L Evans, and A. C Bovik, "Image quality assessment based on a degradation model," *Image Processing, IEEE Transactions on*, vol. 9, no. 4, pp. 636–650, 2002.
- [48] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, no. 3, pp. 247–278, 1998.



**Emille Bosc** received the Engineering degree in electronic and computer engineering from National Institute of Applied Sciences (INSA), Rennes, France, in 2009. She is currently pursuing the Ph.D. degree from INSA, Scientific and Technical University, and in the Institute of Electronics and Telecommunications of Rennes (IETR) Laboratory. Her research interests include 3-D video data representation, 3-D video coding, depth image based rendering on 2-D and 3-D displays, and visual perception understanding in 3-D vision.



quality of medical imaging. He is maintaining subjective assessment facilities for IRCCyN and supervised subjective quality assessment databases releases.

**Romuald PÉpion** received the Engineering degree in electronic and computer engineering from Polytech Nantes in 2005, Nantes, France. He is Research Engineer at Polytech Nantes/Université de Nantes in the Image and Video Communication Group (35 researchers) at CNRS IRCCyN lab. He is mostly engaged in research dealing with image and video quality assessment. He has been involved in several collaborative research programs: HD4U on HDTV, scalimages on H264 SVC, Futurimages on 3-D and HDR quality assessment, SVC4QoE, Equimose on



ence, image processing, and computer vision. Her subject was dealing with projective invariants applied to computer vision. She is a member of the IETR/INSA de Rennes Laboratory. Her research activities deal with 3-D modelization for video sequence communication.

**Luce Morin** received the M.S. degree from ENSPS school in Strasbourg in 1989 and the Ph.D. degree with Prof. R. Mohr, from the LIFIA laboratory, INP-Grenoble, France. After the M.S. degree, she had a 6 month end-of-study internship at the NASA Goddard Space Flight Center, Washington, DC. From 1993 to 2008 she was an associate professor at University of Rennes and a member of the Temics team at IRISA laboratory. She is currently a Full Professor at the National Institute of Applied Sciences (INSA), Rennes, France, where she teaches computer sci-



**Martin Köppel** obtained the degree, Dipl.-Ing. in Media Technologies at the Technical University of Ilmenau, Germany, in 2008. Before then, he worked as a student engineer at the Institut of Microelectronic- and Mechatronic Systems gGmbH, Ilmenau, Germany, from 2004 to 2006. He was a teaching assistant at the Technical University of Ilmenau from 2006 to 2007. He joined the Fraunhofer Institute for Telecommunications Heinrich Hertz Institute, Berlin, Germany, in 2007 and has been working there as a Research Associate since 2008.

His research interests are in the fields of image and video processing. He has been involved in several projects in the areas of texture synthesis, view synthesis, video coding and 3-D video.

**Muriel Pressigout** received the M.S. degree from National Institute of Applied Sciences (INSA) of Rennes, Rennes, France, in 2003 and the Ph.D. degree, in 2006, with Prof. E. Marchand, from the University of Rennes 1, IRISA laboratory, Rennes, France. She is an assistant professor at INSA Rennes. Her thesis work, within the project at IRISA Lagadic focuses on the vision computer and more specifically the real-time monitoring of augmented reality applications and robotics. Her current interests focus on visual servoing, object tracking and 3-D model based image analysis. Her main research theme is 3-D video.



**Patrick Le Callet** received M.Sc. (1997) and PhD (2001) degree in image processing from Ecole polytechnique de l'université de Nantes. He was also student at the Ecole Normale Supérieure de Cachan where he got the Aggregation (credentialing exam) in electronics of the French National Education (1996). He worked as an Assistant professor from 1997 to 1999 and as a full time lecturer from 1999 to 2003 at the department of Electrical engineering of Technical Institute of University of Nantes (IUT). Since 2003, he teaches at Ecole polytechnique de

l'université de Nantes (Engineer School) in the Electrical Engineering and the Computer Science department where he is now Full Professor. Since 2006, he is the head of the Image and Video Communication lab at CNRS IRCCyN, a group of more than 35 researchers. His current centers of interest are color and 3-D image perception, visual attention modeling, video and 3-D quality assessment. He is co-author of more than 150 publications and communications and co-inventor of 14 international patents on these topics. He has coordinated and is currently managing for IRCCyN several National or European collaborative research programs representing grants of more than 3 million Euros. He is serving in VQEG (Video Quality Expert Group) where is co-chairing the "Joint-Effort Group" and "3DTV" activities. He is the French national representative of the European COST action IC1003 QUALINET on Quality of Experience of Multimedia service in which he is leading the working group mechanisms of human perception.



**Patrick Ndjiki-Nya** (M98) received the Dipl.-Ing. title (corr. to M.S degree) from the Technical University of Berlin in 1997. In 2008 he also finished his doctorate at the Technische Universität Berlin. He has developed an efficient method for content-based video coding, which combines signal theory with computer graphics and vision. His approaches are currently being evaluated in equal or similar form by various companies and research institutions in Europe and beyond. From 1997 to 1998 he was significantly involved in the development of a

flight simulation software at Daimler-Benz AG. From 1998 to 2001 he was employed as development engineer at DSPspecialists GmbH where he was concerned with the implementation of algorithms for digital signal processors (DSP). During the same period he researched content-based image and video features at the Fraunhofer Institute for Telecommunications Heinrich Hertz Institute, Berlin, Germany, with the purpose of implementation in DSP solutions from DSPspecialists GmbH. Since 2001 he is solely employed at Fraunhofer Heinrich Hertz Institute, where he was Project Manager initially and Senior Project Manager from 2004 on. He has been appointed group manager in 2010.