# Versatile input view selection for efficient immersive video transmission

Dominika Klóska
dominika.kloska@put.poznan.pl

Adrian Dziembowski
adrian.dziembowski@put.poznan.pl

Jarosław Samelak
jaroslaw.samelak@gmail.com

Institute of Multimedia Telecommunications, Poznań University of Technology
Polanka 3, 61-131 Poznań, Poland

## ABSTRACT

In this paper we deal with the problem of the optimal selection of input views, which are transmitted within an immersive video bitstream. Due to limited bitrate and pixel rate, only a subset of input views available on the encoder side can be fully transmitted to the decoder. Remaining views are – in the simplest approach – omitted or – in the newest immersive video encoding standard (MPEG immersive video, MIV) – pruned in order to remove less important information. Selecting proper views for transmission is crucial in terms of the quality of immersive video system user's experience. In the paper we have analyzed which input views have to be selected for providing the best possible quality of virtual views, independently on the viewport requested by the viewer. Moreover, we have proposed an algorithm, which takes into account a non-uniform probability of user's viewing direction, allowing for the increase of the subjective quality of virtual navigation for omnidirectional content.

## Keywords

Immersive video, virtual view synthesis, MPEG immersive video (MIV)

## 1. INTRODUCTION

A natural consequence of rapidly growing interest in immersive video and virtual reality (VR) is the demand for efficient and versatile immersive media transmission. The virtual reality technology allows the user for immersing into the scene captured by a multicamera system and virtually navigating within it (Fig. 1). Such a navigation may be restricted to several degrees of freedom (DoF). For instance, 3DoF systems allow users to rotate their head around a single pivot point, and 3DoF+ systems additionally support restricted, translational movement of user's head [MPEG19], increasing the quality of experience (QoE) when using the head-mounted display (HMD) devices. The latest, most advanced systems – 6DoF – allow users for free, unrestricted navigation within a scene [MPEG17].

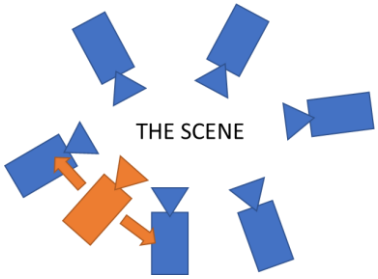


**Figure 1. Idea of an immersive video system; the scene is captured by several cameras (blue), a viewer may virtually change their viewpoint (orange camera).**

In order to obtain an immersive video sequence, it is required to use a multicamera system, containing even hundreds of cameras [Fuj06]. Practical systems contain less cameras (e.g., 10 – 20 [Sta18]), but even in such a case a tremendous amount of data has to be processed and transmitted to the viewer. Moreover, the possibility of virtual immersion into the scene in the immersive video systems is provided by rendering [Fac18], [Sta22] of viewports demanded by the viewer. Such an operation requires information of the three-dimensional scene, which is typically represented in the MVD format (multiview video plus depth, Fig. 2) [Mul18]. Therefore, for each input view also a depth map should be transmitted.

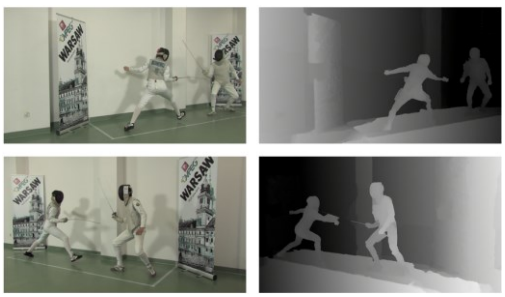

**Figure 2. Sequence in MVD format.**

The easiest way to address the problem of transmission of a huge amount of multiview video data would be to encode each real view (e.g., using HEVC [Sul12]) and the corresponding depth map separately – such an approach is called multiview simulcast. However, this method is not effective due to the high bitrate and pixel rate [Boy21]. Moreover, the quality of the immersive content is not satisfactory because HEVC (or any typical 2D video encoder such as the newest, VVC [Bro21]) encoder was not developed for processing depth maps. It is possible to enhance the quality of the final immersive

vision with the use of MV-HEVC and 3D-HEVC [Tec16], which are HEVC extensions dedicated to encoding 3D content. However, these methods do not guarantee versatility, because they are not adapted for encoding sequences acquired by omnidirectional cameras, or by multicamera systems where the cameras are located arbitrarily. Therefore, none of the abovementioned methods can be used in practical immersive video systems.

The simplest practical solution allowing for a significant decrease of pixel rate is to transmit only a subset of the given real views. In such an approach, in order to obtain the best possible quality, these views have to be carefully chosen.

A more sophisticated, newest approach is based on the use of the MPEG immersive video (MIV) standard [Boy21], [ISO22] which defines the compression of immersive media in a form of multiview video pre- and post-processing combined with the typical video encoder, e.g., VVC [Bro21]. The MIV encoding process can be divided into three main steps, in which the input data ($n$ views and corresponding depth maps) are processed into $k$ video bitstreams called "atlases", further encoded using the VVC encoder (Fig. 3).

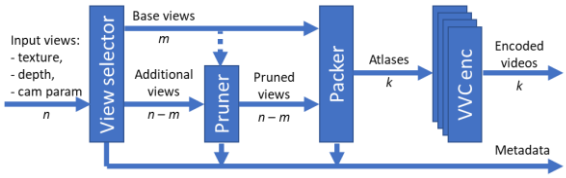


**Figure 3. Simplified scheme of the MIV encoder. Figure from [Dzi22a].**

In the first step the MIV decides which views are the most important from the user's point of view. These views are then being labeled as "base views" and are being placed in atlases in their entirety (Fig. 4A and C). Remaining views ("additional views") contain a lot of redundant data are then pruned in order to remove the excess data. Finally, after the pruning operation additional views are packed into atlases as a form of patch mosaic (Fig. 4B and D) [Vad22].

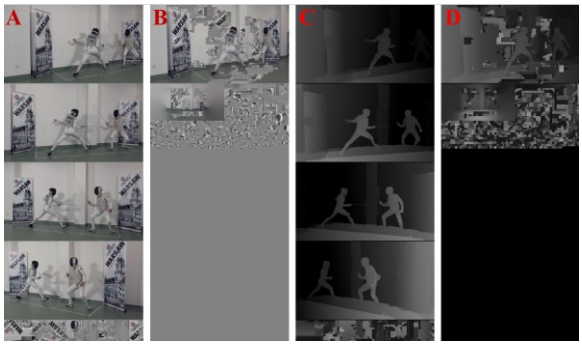


**Figure 4. Four atlases produced by the MIV encoder using the MIV Main profile [Vad22]: 2 texture atlases (A, B) and 2 depth atlases (C, D).**

On the decoder side, the atlases are firstly decoded using the typical video decoder (such as VVC). After the video decoding step, the views and depths stored in atlases are unpacked and then used for rendering of the views requested by user (Fig. 5).

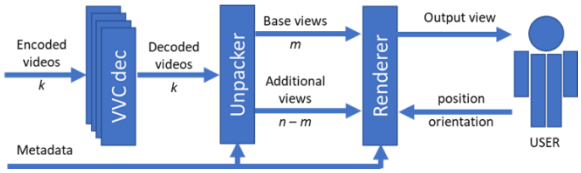


**Figure 5. Simplified scheme of the MIV decoder. Figure from [Dzi22a].**

Irrespectively of the immersive video coding approach, the effective input view selection is the crucial step in terms of providing the best quality and the highest coding efficiency. In this paper, we describe the view selection method which allows for efficient immersive video transmission in practical systems, where it is not possible to send all the real views to the decoder. Moreover, proposed algorithm performs efficiently both for content acquired by perspective and omnidirectional cameras, and can be used for 6DoF systems, where the user virtually immerses into the scene [Laf17].

## 2. INPUT VIEW SELECTION

### 2.1. View selection for virtual view synthesis

The proper input view selection method should provide the highest possible quality of synthesized views while preserving similar bitrate. Considerations on the influence of input view selection on the virtual view quality were described by the authors of this paper in [Dzi18], where we focused on optimizing the quality in simple free navigation systems [Sta18]. In [Dzi18], we assumed that the renderer has access to all of input views, but - in order to provide reasonable computational time - it can use only two of them for rendering purposes.

The input view choice requires addressing three problems: occlusions, finite resolution of video, and non-Lambertian surfaces; leading to the conclusion that the highest quality of rendered views can be obtained based on nearest left and nearest right input view. Obviously, in such a scenario it would be optimal to transmit these two views and skip all the others. However, a selection of these two views is possible only if the position of view requested by the viewer is known before the transmission.

### 2.2. View selection for immersive video transmission

In a practical immersive video system, where multiple viewers receive the same bitstream and are able to independently choose their point of view [Tan12], an assumption regarding viewer's position known *a priori* before the transmission is invalid. Instead, it is required to choose input views in the

way, which guarantees the highest average quality of views watched by users, independently of their viewpoint.

In order to meet the requirements for immersive data transmission, where the position of a user cannot be predicted, the view selection method described in [Dzi18] has to be extended.

Taking into account the statement that the quality of synthesized view is highest when the rendering is performed on the basis of the nearest left and right real view, we have conducted a simulation. In the simulation we assumed a simple practical immersive video system with reasonable pixel rate [Boy21] and number of cameras [Sal18]:

- linear multicamera system with 13 evenly distributed cameras,
- 13 input views available at the encoder side,
- 4 input views transmitted to the decoder,
- 100 possible virtual positions of the viewer (evenly distributed too).

We assumed that the left-most and right-most input views are transmitted (in order to ensure, that for all virtual positions of the viewer there exists left and right input view). Therefore, the index of the first transmitted view was fixed to 1, and index of the fourth view was fixed to 13. Indices of remaining two input views to be transmitted were unknown, and they were iteratively changed in order to calculate their optimal position.

For each virtual position of the viewer, we calculated the total distance to the nearest left and nearest right view. The results are presented in Fig. 6, where the horizontal axis presents the index of the first real view used for virtual view synthesis, the vertical axis presents the index of the second real view.

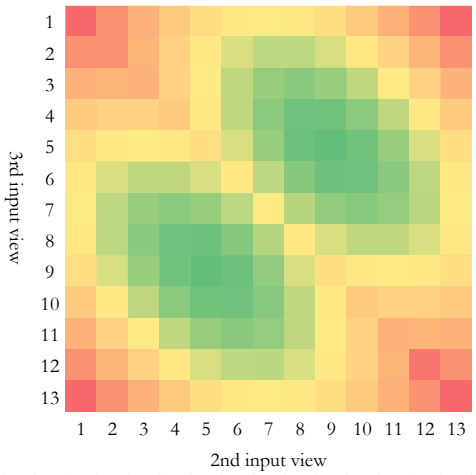


**Figure 6. Total distance between nearest left and nearest right view calculated for all real views used for the experiment.**

The green color in Fig. 6 indicates that the total distance between the virtual view and its transmitted neighbors was low, red – that the distance was higher. These results show that the distance is minimized when the input views selected to be transmitted are distributed evenly. Of course, in Fig. 6 we presented only the distance measured for a simple simulation, not the quality of synthesized views. Therefore, in the next section we presented evidence for these considerations.

It should be noted that the presented model and experiment assumed a simple, linear camera arrangement. However, analogous conclusions can be taken also for more sophisticated multicamera systems.

# 3. EXPERIMENTAL RESULTS

## 3.1. Methodology

In order to prove authenticity of considerations presented in the previous section, we have performed an experiment. In the experiment, we assessed the quality of virtual views synthesized using various combinations of input views. The view synthesis was performed using the MPEG's reference software – VVS [Dzi19].

The test set comprised of two computer-generated sequences – BBB Butterfly and BBB Flowers [Kov15]. We have decided to use these sequences, as they contain multiple (79) input views, making possible quality assessment for several viewpoints. Both sequences were captured by 79 evenly-distributed cameras placed on an arc. For each sequence, views are numbered from v6 to v84. Seven views: v6, v19, v32, v45, v58, v71, and v84 were used as input ones, while all remaining views were treated as reference for objective quality evaluation.

The quality of synthesized views was calculated using two objective quality metrics, described in the MIV Common Test Conditions (MIV CTC) [MPEG22c] and commonly used in the experiments related to immersive video: WS-PSNR [Sun17] and IV-PSNR [Dzi22b]. Both quality metrics are full-reference ones, therefore the quality was assessed by comparing input views with virtual views synthesized in the same position (the same viewpoint).

In the experiment we assumed the transmission of four input views (of the seven available). Two input views were fixed: v6 as the first input view and v84 as the fourth one. The position of second and third input views was being changed in order to define the optimal arrangement of transmitted views.

## 3.2. Results

Mean IV-PSNR and WS-PSNR of synthesized virtual views are presented in Tables 1 and 2. The values were averaged over 75 synthesized views (v6 to v84, excluding four views used as input ones, for

which the quality is perfect, as no synthesis is needed).

| 3rd input view \ 2nd input view | v19 | v32 | v45 | v58 | v71 |
|---|---|---|---|---|---|
| v19 | | 31.32 | 32.06 | 32.32 | 31.34 |
| v32 | 31.32 | | 32.10 | 33.16 | 32.57 |
| v45 | 32.06 | 32.10 | | 31.78 | 31.59 |
| v58 | 32.32 | 33.16 | 31.78 | | 30.74 |
| v71 | 31.34 | 32.57 | 31.59 | 30.74 | |

| 3rd input view \ 2nd input view | v19 | v32 | v45 | v58 | v71 |
|---|---|---|---|---|---|
| v19 | | 39.39 | 41.08 | 41.03 | 39.44 |
| v32 | 39.39 | | 40.78 | 41.65 | 40.36 |
| v45 | 41.08 | 40.78 | | 40.04 | 39.79 |
| v58 | 41.03 | 41.65 | 40.04 | | 38.05 |
| v71 | 39.44 | 40.36 | 39.79 | 38.05 | |

**Table 1. Mean IV-PSNR [dB] of virtual view (averaged over 75 views) calculated for different combinations of transmitted input views. 1st input view was set to v6, 4th input view: v84. Sequences: BBB Flowers (top) and BBB Butterfly (bottom).**

| 3rd input view \ 2nd input view | v19 | v32 | v45 | v58 | v71 |
|---|---|---|---|---|---|
| v19 | | 24.12 | 24.91 | 24.99 | 23.99 |
| v32 | 24.12 | | 24.98 | 25.68 | 25.03 |
| v45 | 24.91 | 24.98 | | 24.64 | 24.42 |
| v58 | 24.99 | 25.68 | 24.64 | | 23.57 |
| v71 | 23.99 | 25.03 | 24.42 | 23.57 | |

| 3rd input view \ 2nd input view | v19 | v32 | v45 | v58 | v71 |
|---|---|---|---|---|---|
| v19 | | 31.99 | 33.05 | 32.91 | 31.80 |
| v32 | 31.99 | | 32.90 | 33.52 | 32.68 |
| v45 | 33.05 | 32.90 | | 32.40 | 32.26 |
| v58 | 32.91 | 33.52 | 32.40 | | 30.91 |
| v71 | 31.80 | 32.68 | 32.26 | 30.91 | |

**Table 2. Mean WS-PSNR [dB] of virtual view (averaged over 75 views) calculated for different combinations of transmitted input views. 1st input view was set to v6, 4th input view: v84. Sequences: BBB Flowers (top) and BBB Butterfly (bottom).**

As presented, in all considered scenarios (both sequences and both quality metrics), the best average quality of synthesized views can be achieved when using views v6, v32, v58, and v84, thus evenly distributed input views.

Such a view selection provides also highest quality in a worst-case scenario (the lowest quality among all synthesized views, Table 3).

Moreover, even distribution of transmitted input views minimizes ΔIV-PSNR (difference between lowest and highest quality among all synthesized views, Table 4), making the user's experience more stable, as the perceived quality change during virtual navigation among the scene is lower.

Tables 3 and 4 present only the results obtained for the IV-PSNR metric. The WS-PSNR results were omitted, as it behaves similarly, and the best results were achieved for evenly distributed input views.

| 3rd input view \ 2nd input view | v19 | v32 | v45 | v58 | v71 |
|---|---|---|---|---|---|
| v19 | | 22.95 | 26.94 | 23.50 | 23.40 |
| v32 | 22.95 | | 26.94 | 29.72 | 25.72 |
| v45 | 26.94 | 26.94 | | 24.41 | 24.17 |
| v58 | 23.50 | 29.72 | 24.41 | | 21.41 |
| v71 | 23.40 | 25.72 | 24.17 | 21.41 | |

| 3rd input view \ 2nd input view | v19 | v32 | v45 | v58 | v71 |
|---|---|---|---|---|---|
| v19 | | 31.54 | 36.28 | 35.08 | 32.28 |
| v32 | 31.54 | | 36.15 | 38.36 | 34.95 |
| v45 | 36.28 | 36.15 | | 33.75 | 33.38 |
| v58 | 35.08 | 38.36 | 33.75 | | 30.38 |
| v71 | 32.28 | 34.95 | 33.38 | 30.38 | |

**Table 3. Lowest IV-PSNR [dB] of virtual view (among 75 views) calculated for different combinations of transmitted input views. 1st input view was set to v6, 4th input view: v84. Sequences: BBB Flowers (top) and BBB Butterfly (bottom).**

| 3rd input view \ 2nd input view | v19 | v32 | v45 | v58 | v71 |
|---|---|---|---|---|---|
| v19 | | 18.99 | 14.99 | 18.74 | 21.14 |
| v32 | 18.99 | | 14.99 | 12.52 | 18.82 |
| v45 | 14.99 | 14.99 | | 17.83 | 20.37 |
| v58 | 18.74 | 12.52 | 17.83 | | 23.15 |
| v71 | 21.14 | 18.82 | 20.37 | 23.15 | |

| 3rd input view \ 2nd input view | v19 | v32 | v45 | v58 | v71 |
|---|---|---|---|---|---|
| v19 | | 14.06 | 9.91 | 10.12 | 12.67 |
| v32 | 14.06 | | 10.16 | 7.83 | 9.94 |
| v45 | 9.91 | 10.16 | | 13.35 | 12.49 |
| v58 | 10.12 | 7.83 | 13.35 | | 16.20 |
| v71 | 12.67 | 9.94 | 12.49 | 16.20 | |

**Table 4. ΔIV-PSNR [dB] of virtual view (among 75 views) calculated for different combinations of transmitted input views. 1st input view was set to v6, 4th input view: v84. Sequences: BBB Flowers (top) and BBB Butterfly (bottom).**

# 4. INPUT VIEW SELECTION FOR OMNIDIRECTIONAL CONTENT

## 4.1. Omnidirectional content problem

All the considerations presented in previous sections assumed that the viewer can watch the scene from any viewpoint, and the probability of choosing various viewpoints is the same. However, it is not true for 6DoF and 3DoF+ [Wie19] immersive video systems, where the user virtually immerses into the scene, e.g., using the HMD device. In such a case, a typical user tends to look around in the horizontal

plane, while not focusing on floor, ceiling or the sky above [Dzi22c].

Therefore, the use of the described view selection algorithm, which chooses input views most distant to each other may result in non-optimal selection. For instance, views captured by cameras facing down or up may be selected instead of views containing essential information about the scene (Fig. 7).


**Figure 7. Three views of the Chess sequence [Ilo19].**

The example presented in Fig. 7. A and B are two of 7 views selected as "base views" by the MIV encoder (working under the decoder-side depth estimation – DSDE – configuration [Mie22]). Fig. 7.C presents a view, which was selected as an "additional view" and skipped despite having more important information from the viewer's perspective.

Such a selection increases the quality of the floor and the ceiling but decreases the quality of a chess knight – which is more crucial for the viewer.

## 4.2. Proposed solution

Taking into account the subjective non-uniform significance of different areas of the scene, we proposed a modification of the simple view selection algorithm, which penalizes the vertical distance between cameras.

In the basic approach (e.g., the one implemented in the 14th version of the Test Model for MPEG immersive video – TMIV 14 [MPEG22a]), basic views were selected by maximization of the total distance between them. The distance between two views *i* and *j* was calculated as:

$$r_{i,j} = \sqrt{\left(r_{i,j}^{x}\right)^2 + \left(r_{i,j}^{y}\right)^2 + \left(r_{i,j}^{z}\right)^2}, \qquad (1)$$

where $r_{i,j}^{x}$, $r_{i,j}^{y}$, and $r_{i,j}^{z}$ are distances between views *i* and *j* along three axes of the global coordinate system.

We proposed to modify (1) by addressing the non-uniform probability of viewer's watching direction and by penalizing the vertical distance. To achieve that the camera distances calculated in the view selection process are not homogenous meaning that the vertical direction is being treated differently from horizontal directions:

$$r_{i,j} = \sqrt{\left(r_{i,j}^{x}\right)^2 + \left(r_{i,j}^{y}\right)^2 + \left(w \cdot r_{i,j}^{z}\right)^2}, \qquad (2)$$

Where $r^x, r^y, r^z$ indicate a distance between two cameras among three axes, and where *w* is the inhomogeneity coefficient. In the experiments described in the further part of this section, the *w* value was set to 0.4.

The proposed change allows for selecting input views, which carry valuable information (Fig. 8.B) instead of sending views containing plain floor or ceiling of the scene, irrelevant for the viewer (Fig. 8).

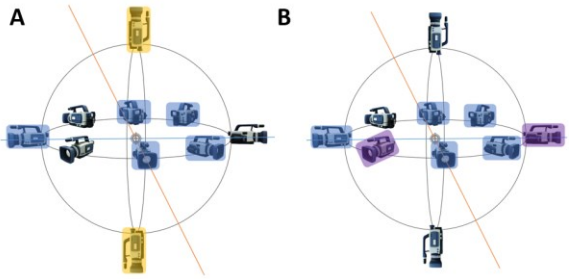


**Figure 8. Sequence Chess. A – views selected for encoding by the anchor method (blue color highlights selected cameras that carry valuable information about the scene, yellow shows selected cameras that have less importance to the viewer and therefore can be omitted), B – views selected for encoding after the modification (purple color highlights cameras that were selected instead of the yellow cameras from Fig 8.A).**

Cameras highlighted in blue were selected as base views for both basic and modified view selection algorithms. Besides them, the basic algorithm selected cameras facing up and down (yellow cameras in Fig. 8), while the modified algorithm – two cameras acquiring important parts of the scene. Considering the fact, that a typical viewer spends more time looking around on the horizontal plane rather than the vertical one (which contains the floor and the ceiling) [Dzi22c], we propose to send more views from the horizontal plane instead of the views facing upwards and downwards. It will have a positive influence on the final quality of the particular parts of the scene at which the user looks the majority of the time.

The proposed modification was appreciated by the experts of the ISO/IEC JTC1/SC29/WG 04 MPEG VC group and is included in the newest version of the Test Model for MIV – TMIV 15 [MPEG22b].

The influence of this view selection modification on the objective and subjective quality of the final immersive vision was described in the following subsections.

## 4.3. Methodology of the experiment

The experiment was conducted under the common test conditions for MPEG immersive video (MIV CTC) [MPEG22c], but the test set was limited to omnidirectional sequences only (fig 9).

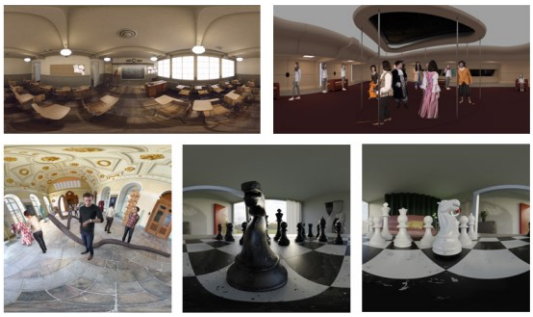

**Figure 9. Omnidirectional sequences used for the experimental results, sequences: ClassroomVideo [Kro18], Hijack [Dor18], Museum [Dor18], Chess [Ilo19], and ChessPieces [Ilo20].**

In the experiment, the TMIV 14 software [MPEG22a] was used. Each sequence was compressed with the use of five different rate points (RP) using the VVC encoder. The total bitrate for a sequence ranged from 2.5 Mbps (RP5) to over 20 Mbps (RP1). The quality of the synthesized virtual views was assessed using WS-PSNR and IV-PSNR objective quality metrics [Sun17], [Dzi22b].

In order to perform a thorough test, three configurations of TMIV were evaluated: MIV, MIV View and MIV DSDE. The detailed description of these configurations can be found in the publicly available MIV CTC document [MPEG22c].

Besides the objective quality measurement, also the subjective quality of rendered views was evaluated. The subjective quality assessment was performed based on pose traces [Boy21], according to the MIV CTC.

The subjective quality evaluation was done by 45 naïve viewers, watching two side-by-side videos (Pair Comparison method, Rec. ITU-T P.910 [ITU08] and ITU-R BT.500 [ITU98]). The viewers judged the quality of presented posetraces with the use of 7-number scale with values from -3 to 3.

To minimize the time duration of the subjective test, only three RP were shown to the viewers: RP1, RP3, and RP5. Moreover, subjects were assessing quality change only for sequences, for which the view selection result was different, than for unmodified TMIV14 (see Table 5).

| Sequence Name \ MIV Configuration | MIV Main | MIV View | MIV DSDE |
|---|---|---|---|
| Classroom | × | × | × |
| Museum | ✓ | ✓ | ✓ |
| Hijack | × | × | ✓ |
| Chess | × | × | ✓ |
| ChessPieces | × | × | ✓ |

**Table 5. Overview of the sequences and different MIV configurations. „X" indicates the scenario in which view selection result was the same as with the unmodified TMIV 14 software.**

**a) Subjective quality evaluation**

The results of performed subjective quality evaluation are presented in Figs. 10 and 11. In Fig. 10, an influence of the proposed method on efficiency of different MIV configurations are presented. Fig. 11 contains comparison of subjective quality change for different test sequences. The results are presented as an average quality change caused by the proposed modification and the 95% confidence interval, calculated according to ITU-R recommendations [ITU98] as:

$$CI = 1.96 \cdot \frac{SD}{\sqrt{N}}, \qquad (3)$$

where CI is the confidence interval, SD – standard deviation, and N – number of viewers (in presented experiment N = 45).

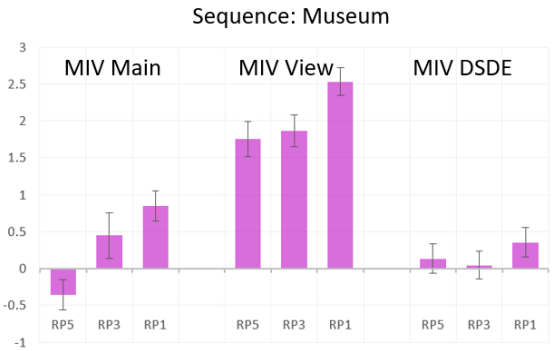


**Figure 10. Subjective results for the Museum sequence in three different MIV configurations.**

Subjective quality changes presented for the Museum sequence in Fig. 10 show that in almost every scenario there was a visible quality improvement. For 6 of 9 tests, the proposal allowed for achieving a statistically important quality improvement. For two tests (RP5 and RP3 in MIV DSDE configuration) the quality gain was also spotted, but it was not statistically important.

The proposal decreased the subjective quality in only one case – the heaviest compression in the MIV Main scenario. However, as presented in Fig. 12, for such a low bitrate the MIV Main cannot properly handle Museum sequence, and the quality of the content was unsatisfactory also before proposed modification.

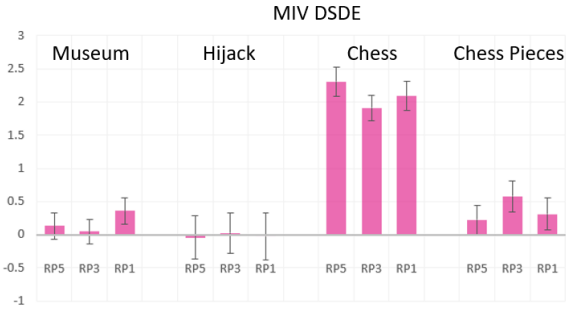


**Figure 11. Subjective results for four sequences in MIV DSDE configuration.**

Analysis of the results obtained for different sequences in the MIV DSDE configuration (Fig. 11) is similar to the results presented in Fig. 10. For 3 of 4 test sequences there is a quality gain induced by the proposed view selection modification. Moreover, for these sequences this gain is statistically important in 7 of 9 cases.

The only sequence for which the quality change did not occur is Hijack, but it should be noted that the proposal did not decrease the subjective quality.

**b) Objective results**

Objective quality results are gathered in Tables BDR and BDP, and presented as Bjøntegaard deltas [Bjo01]: BD-rate (Table 6) and BD-PSNR (Table 7).

| Sequence | | High-BR BD rate WS-PSNR | Low-BR BD rate WS-PSNR | High-BR BD rate IV-PSNR | Low-BR BD rate IV-PSNR |
|---|---|---|---|---|---|
| Museum | MIV | -3.4% | 0.2% | 0.9% | 2.3% |
| | MIV View | -4.6% | -5.5% | -7.7% | -3.2% |
| | MIV DSDE | 57.2% | 32.5% | 13.1% | 10.2% |
| Hijack | | --- | --- | --- | --- |
| Chess | | --- | --- | --- | --- |
| ChessPieces | | --- | --- | --- | --- |

**Table 6. Objective metric (WS-PSNR and IV-PSNR) BD-rates obtained for 4 lowest rate points (Low-BR) and for 4 highest rate points (High-BR); "---" denotes, that the BD-rate calculation was not possible because of non-overlapping curves.**

| Sequence | | High-BR BD rate WS-PSNR | Low-BR BD rate WS-PSNR | High-BR BD rate IV-PSNR | Low-BR BD rate IV-PSNR |
|---|---|---|---|---|---|
| Museum | MIV | 0.4% | 0.1% | -0.0% | -0.2% |
| | MIV View | 0.2% | 0.3% | 0.8% | 0.3% |
| | MIV DSDE | -1.6% | -1.3% | -0.8% | -0.6% |
| Hijack | | -13.2% | -13.2% | -11.1% | -11.1% |
| Chess | | -6.4% | -6.2% | -5.9% | -5.5% |
| ChessPieces | | -11.3% | -11.0% | -8.3% | -7.9% |

**Table 7. Objective metric (WS-PSNR and IV-PSNR) BD-PSNRs obtained for 4 lowest rate points (Low-BR) and for 4 highest rate points (High-BR).**

Surprisingly, presented objective results show, that in general the proposed method performs worse than the basic view selection algorithm implemented in TMIV 14. However, it has to be highlighted that results presented in Tables 6 and 7 were obtained by averaging the IV-PSNR and WS-PSNR values of all synthesized views, including the basic views. An example is shown in Table 8, where exact IV-PSNR values for all 10 views of sequence Chess are presented.

As presented in Table 8, the average IV-PSNR for non-base (i.e., "additional") views is similar to the quality obtained for unmodified TMIV14. The only views with significant quality degradation are v0 and v9 (i.e., views captured by cameras facing up and down), which are less important to the viewer.

| | IV-PSNR [dB] | | |
|---|---|---|---|
| View | TMIV 14 | Proposed | delta |
| v0 | 57.00 | 36.17 | -20.84 |
| v1 | 37.84 | 37.99 | 0.15 |
| v2 | 56.78 | 56.76 | -0.02 |
| v3 | 33.20 | 56.28 | 23.09 |
| v4 | 56.73 | 56.44 | -0.29 |
| v5 | 56.01 | 55.93 | -0.08 |
| v6 | 38.85 | 56.11 | 17.26 |
| v7 | 56.02 | 56.04 | 0.03 |
| v8 | 57.08 | 57.18 | 0.10 |
| v9 | 56.34 | 30.08 | -26.26 |
| **Average (all views)** | | | **-0.69** |
| **Average (only non-base views)** | | | **-0.02** |

**Table 8. IV-PSNR of synthesized views, RP1, similar bitrate for both approaches (21.3 Mbps for TMIV 14 and 20.8 Mbps for proposed); Chess sequence, MIV DSDE configuration.**

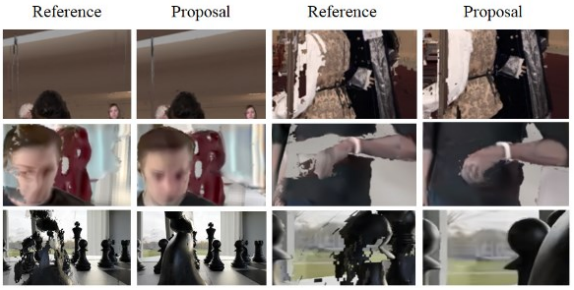


**Figure 12. Visual comparison of posetraces generated with the use of original (reference) and proposed view selection method; sequences (from top): Hijack, Museum, and Chess.**

## 5. CONCLUSIONS

This paper deals with problems on view selection for immersive video and its influence on the quality of final immersive vision on the decoder side.

Firstly, we have conducted an experiment to assess which views from multiview sequences should be selected into the virtual view synthesis process in order to obtain the best quality possible. Received results proved that the view selection algorithm should select views that are evenly distributed.

Moreover, in the paper we proposed an algorithm, which increases the subjective quality of virtual navigation by taking into account a non-uniform probability of choosing the viewing direction. We have noticed, that a typical user usually chooses to watch the scene in the horizontal plane, while the top and bottom parts of the omnidirectional scene are less important. This proposal was appreciated by the ISO/IEC MPEG VC experts, and is included in the reference software [MPEG22b] for the MPEG immersive video coding standard [ISO22].

The proposed approach is based on observations on the behavior of a typical user, without thorough

statistical analysis. Moreover, a correlation between optimal view selection and scene characteristics should be taken into consideration. Therefore, the topic of view selection for immersive video transmission will be studied further in our future research.

## 6. ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science of Republic of Poland.

## 7. REFERENCES

[Bjo01] Bjøntegaard, G. Calculation of average PSNR differences between RD986 curves. ISO/IEC JTC1/SC29/WG11 MPEG M15378, Austin, TX, 2001.

[Bro21] Bross B. et al. Overview of the Versatile Video Coding (VVC) standard and its applications. IEEE T. on Circuits and Systems for Video Technology 31 (10), pp. 3736-3764, 2021.

[Boy21] Boyce J. et al. MPEG Immersive Video coding standard. Proceedings of the IEEE 109 (9), pp. 1521-1536, 2021.

[Dor18] Doré R. Technicolor 3DoF+ Test Materials. Doc. ISO/IEC JTC1/SC29/WG11 MPEG/ M42349, 2018.

[Dzi18] Dziembowski A. et al. View selection for virtual view synthesis in free navigation systems. International Conference on Signals and Electronic Systems 2018, Kraków, Poland, 2018.

[Dzi19] Dziembowski A. et al. Virtual view synthesis for 3DoF+ video. Picture Coding Symposium, PCS 2019, 2019.

[Dzi22a] Dziembowski A. et al. Spatiotemporal redundancy removal in immersive video coding, Journal of WSCG, vol. 30, no. 1-2, pp. 54-62, 2022.

[Dzi22b] Dziembowski A. et al. IV-PSNR – the objective quality metric for immersive video applications. IEEE T. on Circuits and Systems for Video Technology 32 (11), pp. 7575-7591, 2022.

[Dzi22c] Dziembowski, A., Klóska, D., Jeong, J.Y., and Lee, G. [MIV] Inhomogeneous view selection for omnidirectional content. ISO/IEC JTC1/SC29/WG4 MPEG 140, M60668, 10.2022.

[Fac18] Fachada S. et al. Depth image based view synthesis with multiple reference views for virtual reality. 3DTV-Conf, 2018.

[Fuj06] Fujii T. et al. Multipoint measuring system for video and sound – 100-camera and microphone system, ICME conference, 2006.

[Ilo19] Ilola L. et al. New test content for immersive video – Nokia Chess. Doc. ISO/IEC JTC1/SC29/WG11 MPEG, M50787, 2019.

[Ilo20] Ilola L., Vadakital V.K.M. Improved NokiaChess sequence. Doc. ISO/IEC JTC1/SC29/WG11 MPEG, M57382, 2020.

[ISO22] Standard ISO/IEC FDIS 23090-12. Information technology – Coded representation of immersive media – Part 12: MPEG Immersive video. 2022.

[ITU98] Methodology for the Subjective Assessment of the Quality of Television Pictures, document Rec. ITU-R BT.500-9, ITU-R, 1998.

[ITU08] Subjective Video Quality Assessment Methods for Multimedia Applications, document Rec. ITU-T P.910, ITU-T, 2008.

[Kov15] Kovacs, P. BBB light-field test sequences. ISO/IEC JTC1/SC29/WG11, M35721, 2015.

[Kro18] Kroon B. 3DoF+ test sequence ClassroomVideo, ISO/IEC JTC1/SC29/WG11 MPEG M42415, 2018.

[Mie22] Mieloch D. et al. Overview and Efficiency of Decoder-Side Depth Estimation in MPEG Immersive Video. IEEE T. on Circ. and Systems for Video Tech. 32 (9), pp. 6360-6374, 2022.

[MPEG17] Requirements on 6DoF (v1), ISO/IEC JTC1/SC29/WG11 MPEG N17073, 2017.

[MPEG19] MPEG Call for Proposals on 3DoF+ Visual, ISO/IEC JTC1/SC29/WG11 MPEG/N18145 2019.

[MPEG22a] Test Model 14 for MPEG immersive video, ISO/IEC JTC1/SC29/WG04 MPEG VC N0242, 2022.

[MPEG22b] Test Model 15 for MPEG immersive video. Document ISO/IEC JTC1/SC29/WG04 MPEG VC, N0271, 2022.

[MPEG22c] Common test conditions for MPEG immersive video. ISO/IEC JTC1/SC29/WG04 MPEG VC, N0232, 2022.

[Mul18] Müller K. et al. 3-D Video Representation Using Depth Maps. Proceedings of the IEEE 99 (4), pp. 643-656, 2011.

[Sal18] Salahieh, B. et al. Kermit test sequence for Windowed 6DoF activities. ISO/IEC JTC1/SC29/WG11 MPEG M43748, Ljubljana, Slovenia, 2018.

[Sta18] Stankiewicz O. et al. A free-viewpoint television system for horizontal virtual navigation. IEEE Transactions on Multimedia 20 (8), pp. 2182-2195, 2018.

[Sta22] Stankowski J., Dziembowski A. Real-time CPU-based view synthesis for omnidirectional video. 30. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, WSCG 2022, 2022.

[Sul12] Sullivan G. et al. Overview of the High Efficiency Video Coding (HEVC) standard. IEEE Transactions on Circuits and Systems for Video Technology 22 (12), pp. 1649-1668, 2012.

[Sun17] Sun Y. et al. Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video. IEEE Signal Processing Letters 24 (9), pp. 1408-1412, 2017.

[Tan12] Tanimoto M. et al. FTV for 3-D Spatial Communication. Proceedings of the IEEE, vol. 100, no. 4, pp. 905-917, 2012.

[Tec16] Tech G. et al. Overview of the Multiview and 3D Extensions of High Efficiency Video Coding. IEEE T. on Circuits and Systems for Video Technology 26 (1), pp. 35-49, 2016.

[Vad22] Vadakital V.K.M. et al. The MPEG immersive video standard – current status and future outlook. IEEE MultiMedia 29 (3), 2022.

[Wie19] Wien M. et al., Standardization Status of Immersive Video Coding, IEEE J. on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 5-17, 2019.