High Frame-Rate Virtual View Synthesis Based on Low Frame-Rate Input

Krzysztof Wegner
Institute of Multimedia Telecommunications
Poznan University of Technology
Poznań, Poland
krzysztof.wegner@put.poznan.pl

Jakub Stankowski
Institute of Multimedia Telecommunications
Poznan University of Technology
Poznań, Poland
jakub.stankowski@put.poznan.pl

Olgierd Stankiewicz
Institute of Multimedia Telecommunications
Poznan University of Technology
Poznań, Poland
olgierd.stankiewicz@put.poznan.pl

Hubert Żabiński
Institute of Multimedia Telecommunications
Poznan University of Technology
Poznań, Poland

Krzysztof Klimaszewski
Institute of Multimedia Telecommunications
Poznan University of Technology
Poznań, Poland
krzysztof.klimaszewski@put.poznan.pl

Tomasz Grajek
Institute of Multimedia Telecommunications
Poznan University of Technology
Poznań, Poland
tomasz.grajek@put.poznan.pl

***Abstract*— In the paper we investigated the methods of obtaining high-resolution, high frame-rate virtual views based on low frame-rate cameras for applications in high-performance multiview systems. We demonstrated how to set up synchronization for multiview acquisition systems to record required data and then how to process the data to create virtual views at a higher frame rate, while preserving high resolution of the views. We analyzed various ways to combine time frame interpolation with an alternative side-view synthesis technique which allows us to create a required high frame-rate video of a virtual viewpoint. The results prove that the proposed methods are capable of delivering the expected high-quality, high-resolution and high frame-rate virtual views.**

Keywords*— *high-speed camera, multiview acquisition, virtual view synthesis.

I. INTRODUCTION

For years, view synthesis techniques have been developed for the purpose of Free-viewpoint TeleVision (FTV) systems [16, 17]. They allow the generation of additional images of a scene from different positions in space, even from a place where no real camera is present. This gives viewers the freedom of choosing their own viewpoint of a scene, from which they would like to observe the action of the movie. At any moment they can change the viewpoint to any other location within or around a scene.

In order to provide such a functionality, the required images of a scene, called virtual views, are commonly generated by means of view synthesis. The most popular view synthesis technique is Depth Image Based Rendering (DIBR), which uses depth maps to project points/pixels from one view to the other [4, 5, 6, 17].

Apart from the images from real cameras recording a scene, such an algorithm requires additional data, conveying information about the three-dimensional structure of a scene, most commonly provided in the form of depth maps, i.e., greyscale images providing information about the distance between the camera and the objects in a scene. Depth data provided in such a form can be obtained directly via specialized depth cameras like Time of Flight (ToF) cameras [12, 14, 15] or indirectly by estimating depth, based on the analysis of the relative displacement of objects in the captured color images (video) [10, 13] or even by a fusion of these two approaches [7, 9].

In FTV systems, images are captured by a multicamera system composed of a precisely synchronized set of cameras, acquiring images exactly at the same time instance (Fig. 1).

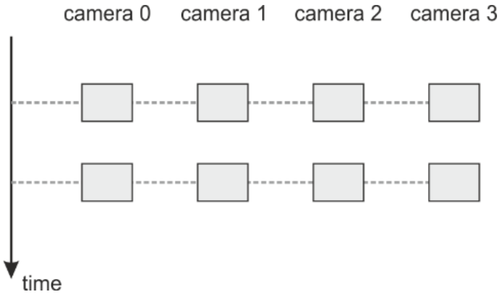


Fig. 1. Synchronous frame acquisition

The frame rate of the cameras used in such systems is usually between 25 to 60 frames per second for high-definition (HD) video.

Of course, there are also higher speed popular cameras available (up to approximately 1000 frames per second), commonly used for "slow motion" effects in handheld devices. Those low-profile high-speed cameras have one significant drawback – they provide the highest frame rate recording only for a limited resolution, and sometimes, for a significantly limited recording period. The use of such cameras in multicamera setups is less common, not only because of the price. Apart from the cost, even at a lower resolution, the use of a high frame-rate camera may produce prohibitively large amounts of data. The use of such cameras in multiview setups is therefore impractical, and may even be undesired, due to the lower resolution of images from high frame-rate cameras.

At the same time, the potential ability of obtaining a high frame-rate, high-resolution multiview recording would undoubtedly expand the capabilities of multiview systems, providing new areas of application. As it was already mentioned, however, such a system, built in a regular way of constructing a multiview system, would be very expensive, and the amounts of data that need to be received and stored, might become overwhelmingly unmanageable.

DOI: 10.1109/MMSP48831.2020.9287076

The purpose of our work is to enable the creation of virtual views at a significantly higher frame rate than it is possible using contemporary solutions, while keeping the cost of the acquisition equipment low and not requiring too much processing power and storage space. In this paper we report our experiments with a mixed domain interpolation, an extension of the already known time interpolation methods. We explore the possible solutions and estimate their performance.

II. State of the Art Methods

The easiest method of obtaining a high frame-rate video is to record it with a high frame-rate camera. But it is also possible to create a high frame-rate video from a normal rate video, albeit acquiring a high frame-rate single view video using low frame-rate camera or cameras is a challenging task. Possibly the easiest is the time interpolation of a low frame-rate video. In such an approach, additional intermediate frames are interpolated based on the captured ones [1, 19] by dragging the objects on their estimated motion trajectories. Nowadays, this process is commonly used in TV monitors to enhance the quality of displayed video [18].

Different methods are also investigated. In the literature, several other ways have been presented to create high frame-rate videos at a lower price than by using specialized cameras, with the use of regular frame-rate cameras.

One of the most interesting options is the use of a multicamera system to create a high-speed video by precisely controlling the acquisition time of each camera in a set (Fig. 2), ensuring a certain delay between triggering several cameras. The cameras are positioned very close to each other, so the recorded images are as similar as possible. After the acquisition, the frame rate can be increased above the frame rate of the cameras used, by switching between the views [20].

The above-mentioned approach has many drawbacks. One of the most severe is that the high-speed video thus obtained is not still, but due to constant view switching, a characteristic shaking at a specific rate can be noticed. Some sophisticated methods can be engineered to combat this shaking, but they may require the use of a complex optical system to split the light coming from the recorded scene between cameras. This method, however, negates the ease of use and significantly increases the cost of such a system. And then, the system is only able to record a single view.

On the other hand, more cost-effective and easy-to-use methods seem to exist that not only enable the free viewpoint capability, but also provide high frame-rate sequences.

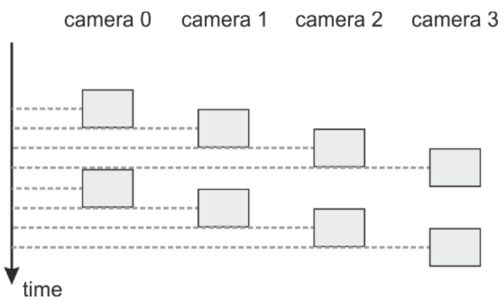


Fig. 2. Time-shifted frame acquisition resulting in a high frame rate

In this paper, we will demonstrate that it is possible to adapt a standard frame-rate multicamera system to create a stable, high frame-rate virtual viewpoint video.

III. Description of the Examined Solutions

There are various ways to create high frame-rate virtual views with the use of a standard frame-rate multicamera acquisition system. The proposed possible approaches are listed below.

A. Synchronized acquisition

In this approach, we use a multicamera system as it is normally designed to be used. We just acquire videos at normal speed, by all cameras, exactly at the same time instance. Then, based on the captured videos, we can synthesize the virtual view with a high frame rate by means of two mechanisms.

1) Virtual view time interpolation

Conceptually the simplest approach (Fig. 3) is to synthesize the virtual view ((1) in Fig. 3) from the input stereo pair and then interpolate ((2) in Fig.3) this virtual view to a higher frame rate. In such an approach, we have a low frame-rate virtual view which is then interpolated to a higher frame rate using the already mentioned object-dragging mechanism.

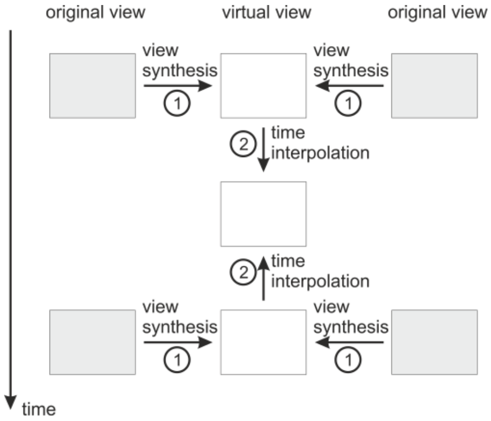


Fig. 3. Obtaining a high frame-rate virtual view by view synthesis first and then time interpolation. (1) – view synthesis, (2) – time interpolation of synthesized views

2) Input view time interpolation

The second approach (Fig. 4) is to first time-interpolate the input views ((1) in Fig. 4) to a higher frame rate and then synthesize the virtual view ((2) in Fig. 4). This approach is

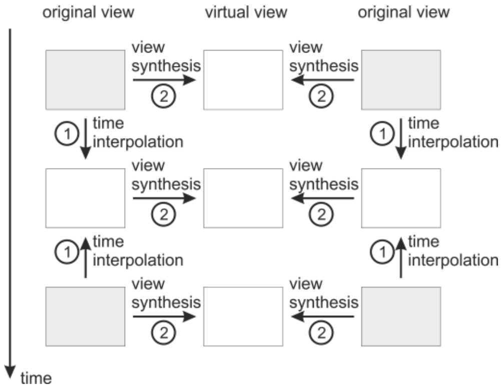


Fig. 4. Obtaining a high frame-rate virtual view by first-time interpolation of the source and then view synthesis. (1) – time interpolation of real camera views and depth maps, (2) – view synthesis

more difficult, as it also requires the interpolation of the depth data along the color images. But as in many applications, one can treat depth data as normal video and use classic video time interpolation tools to add interpolated frames in between the actually recorded frames.

B. *Time-shifted acquisition*

In this approach, we use the multicamera system in a more clever way. Because most multicamera systems employ a very precise synchronization mechanism to control the acquisition moment of all cameras, we can modify this mechanism and shift the acquisition times of, for example, every second camera by half of the acquisition period. This way, odd cameras would record images at odd time instances at normal speed and even cameras would record images at even time instances. Both sets of cameras still acquire video at a normal frame rate, but the entire system records images at double speed, with images at consecutive time instances interleaved between even and odd cameras. Having such time-shifted acquired sequences of views, we can render virtual views at a higher frame rate in several ways.

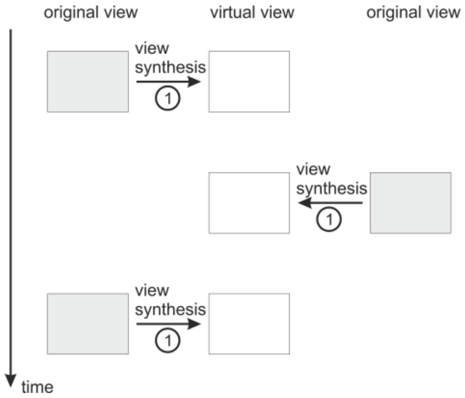


Fig. 5. Obtaining a high frame-rate virtual view by alternating view synthesis. (1) – view synthesis

1) *Alternating virtual view rendering*

The simplest approach would be to take images at an appropriate time instance and project them by view synthesis onto the virtual view position. Every frame of the rendered virtual view will be based on an image from a different camera. One will be based on an image from an odd one, and the next one on an image from an even camera. Each time the virtual view is rendered only from one input view (in general, half of the available captured views) (Fig. 5), which may reduce the quality of the rendered view in some situations. In a general case, not considered in our experiments up to date, for positions in the middle of a large recording rig, the synthesized view in such a scenario may be synthesized from two images – one on the left and one on the right, keeping in mind that the distance between the real and virtual views may be significantly increased at one side (Fig. 6), thus limiting the benefits of two-sided synthesis.

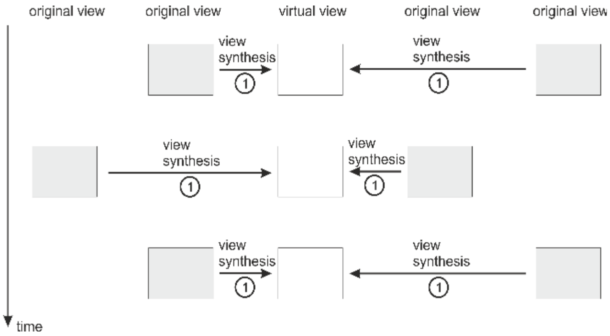


Fig. 6. Obtaining a high frame-rate virtual view by alternating view synthesis for larger camera systems

2) *Input view time interpolation*

Another approach is similar to the one presented in subsection A.2 and exploits time interpolation first (Fig. 7). First, we can interpolate the odd and even camera videos to a higher frame rate ((1) on Fig. 7) and then render the virtual view based on those high frame rate inputs ((2) on Fig. 7). In such an approach, each frame of the virtual view is rendered based on two input views, but those images have different characteristics: one is a frame originally acquired by a real camera and the second one is a time interpolated frame.

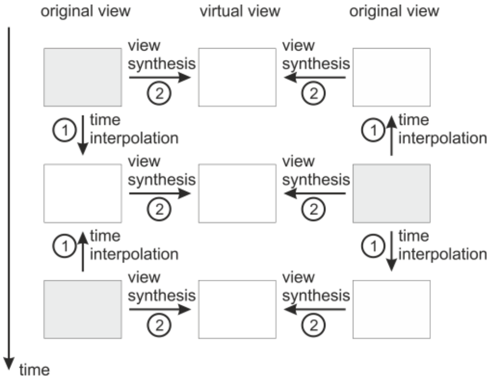


Fig. 7. Obtaining a high frame-rate virtual view by first time interpolation of the source and then view synthesis. (1) – time interpolation of real camera views and depth maps, (2) – view synthesis

IV. Performance of the Described Methods

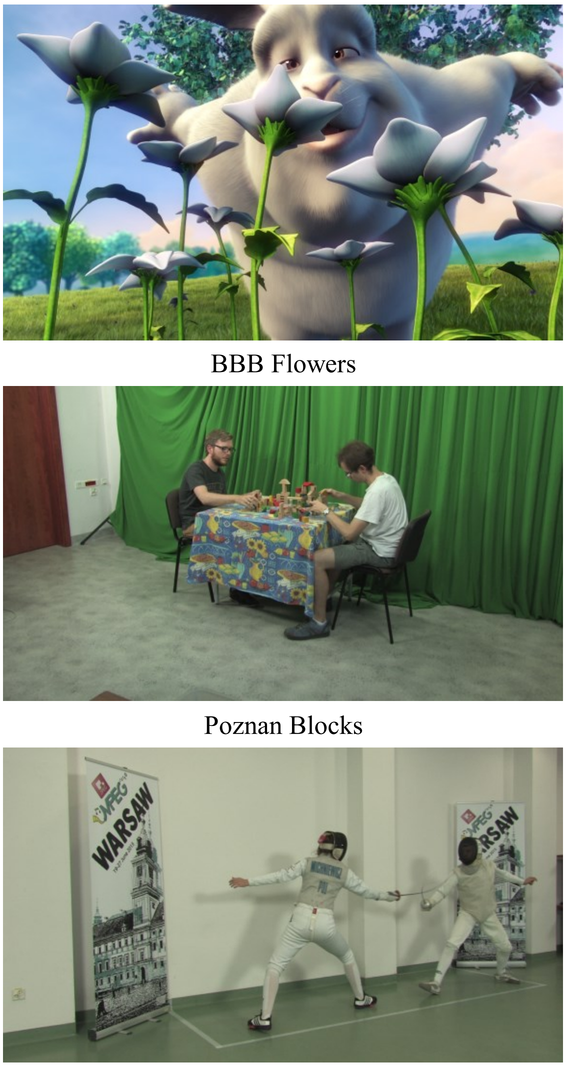
In order to verify the proposed approaches of generation of high frame-rate virtual views using time-shifted image acquisition, we performed the following experiments.

To verify whether the described methods would allow the rendering of high-quality high frame-rate virtual views, one would need a high-speed multiview camera acquisition system to provide the ground truth sequences. Since we do not have such a system at our disposal, and, unfortunately, no such multiview test sequences are publicly available, we utilized a different approach and created some low frame-rate sequences out of normal ones by decimation and then tried to restore the nominal frame rate.

One may argue whether such an approach is a correct one. We acknowledge the fact that the performance of any imaginable time interpolation mechanism would drop in such a situation due to the larger movement of objects between frames for real life sequences. Still, we justify our methodology by the fact that for every configuration in our tests we use the same sequences and the same time interpolation algorithm, thus introducing the same degradation of quality for every case.

We used three multiview test sequences recommended by the MPEG community affiliated by the International Organization for Standardization (ISO). All of the selected test sequences are provided with high-quality depth maps. The depth maps can either be extracted from the scene definition directly – for artificial sequences, or, for real-life sequences, can be calculated using the depth estimation software. In our work we assume that the video data comes with the depth maps of sufficiently good quality.

The sequences used in our experiment had been recorded by a precisely synchronized camera system composed of 5 to 10 cameras at the rate of 25-30 frames per second (depending on a sequence). Out of those sequences we selected only three views. Two of them are used to render a virtual view at the spatial position of the third one. The third view is located exactly in between the selected two others, and it is used as a reference (ground truth) view for quality assessment of the rendered virtual view. A summary of the used test sequences can be found in Table I. Additonally, a single frame from one view from each of the test sequences is presented in Fig. 8.


BBB Flowers

Poznan Blocks

Poznan Fencing

Fig. 8. Examples of frames from the test sequences used

From the selected videos we dropped every other frame to simulate low frame-rate sequences. Depending on the considered scenario, we dropped frames synchronically in two selected input views (simulation of synchronous low frame-rate acquisition) or we dropped frames alternately in the two selected input views (simulation of time-shifted low frame-rate acquisition).

TABLE I. SUMMARY OF THE TEST SEQUENCES USED IN EXPERIMENTS

Sequence name	Views used for rendering		Third view used for quality assessment
BBB Flowers [8]	6	32	19
Poznan Blocks [2]	0	2	1
Poznan Fencing [Dom16]	0	2	1

To the prepared low frame-rate sequences we applied four methods considered in Section III to create a high frame-rate virtual view.

For time interpolation we used open source MVtools2 library used with Avisynth software [11], and for virtual view generation we used a state-of-the-art view synthesizer implemented in the View Synthesis Reference Software (VSRS) [3] package developed by the MPEG community.

For quality assessment we used a widely popular PSNR metric. Each rendered high frame-rate virtual view luminance was compared to the ground truth – luminance of the view at the same spatial location captured by the real camera.

V. RESULTS

Table II presents the average quality of the created high frame-rate video for each test sequence for all test cases. As we can see, the worst results are obtained with the alternating virtual view rendering approach (sec. III.B.1). This is caused by poor virtual view synthesis quality from one reference view. This can be alleviated to some extent in systems with more real cameras available for view synthesis at the expense of more computational power required. The remaining approaches provide very similar results, regardless of the applied order of time interpolation and view synthesis processes.

TABLE II. QUALITY OF THE HIGH FRAME RATE VIRTUAL VIEW MEASURED AS LUMINANCE PSNR AGAINST REAL CAPTURED CAMERA VIEW AT THE SAME SPATIAL LOCATION

Sequence name	Synchronous acquisition		Time-shifted acquisition	
	Virtual view time interpolation (III.A.1) [dB]	*Input view time interpolation (III.A.2) [dB]*	*Alternating virtual view rendering (III.B.1) [dB]*	*Input view time interpolation (III.B.2) [dB]*
BBB Flowers	22.74	22.51	18.41	22.54
Poznan Block	33.03	32.92	26.95	32.92
Poznan Fencing	30.03	30.06	28.00	30.06

Subjective tests performed on a group of 5 experts confirm the objective results – the quality of the high frame-rate sequences is similar in 3 of the 4 tested methods, with the method described in III.B.1 being significantly worse. Example frames for the examined methods are presented in the Fig. 10, Fig 11 and Fig. 12. One image for synchronized acquisition scenario is shown (III.A.2), and one for time-shifted scenario (III.B.1). The (III.B.1) case is the one with the worst quality provided consistently for all examined sequences.

Although, the values of the PSNR metric obtained for the studied cases are low and the expected subjective quality is also low, the results should be viewed in a correct perspective. For our studies on generating a high frame-rate virtual view, we choose the anchor to be the virtual view obtained from the original non-decimated sequence (i.e., no time interpolation is necessary), as shown in Fig. 9. In such a case, the obtained virtual view quality in terms of the PSNR metric is shown in Table III.

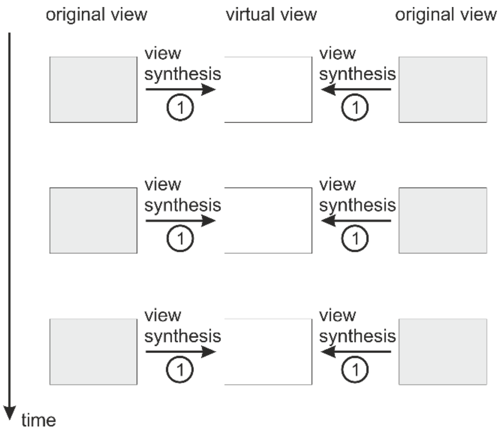


Fig. 9. Obtaining anchor virtual view from original high frame-rate data

TABLE III. QUALITY OF THE ANCHOR VIRTUAL VIEW MEASURED AS LUMINANCE PSNR [DB] AGAINST REAL CAPTURED CAMERA VIEW AT THE SAME SPATIAL LOCATION

Sequence name	Anchor virtual view quality
BBB Flowers	22.79 dB
Poznan Block	32.96 dB
Poznan Fencing	30.33 dB

As it can be seen, the objective quality of the synthesized views from full frame-rate sequences are practically the same as for the cases with time interpolation. Differences between the anchor and scenarios III.A.1, III.A.2 and III.B.2 are a fraction of a decibel. For some cases, the quality of the time-interpolated version outperforms the anchor (Poznan Block sequence for case III.A.1). We expect that the differences would be even smaller if similar experiments were performed with high frame-rate cameras, since the time interpolation for higher frame rates is usually able to provide better results.

Therefore, we feel we can conclude our research by stating that the virtual views generated for the cases with time interpolation do not offer significantly inferior quality to the cases where no time interpolation is necessary.

Our results show that regular frame-rate camera systems can be used to generate interpolated high frame-rate virtual views with no significant degradation of virtual view quality, at a significantly lower cost.

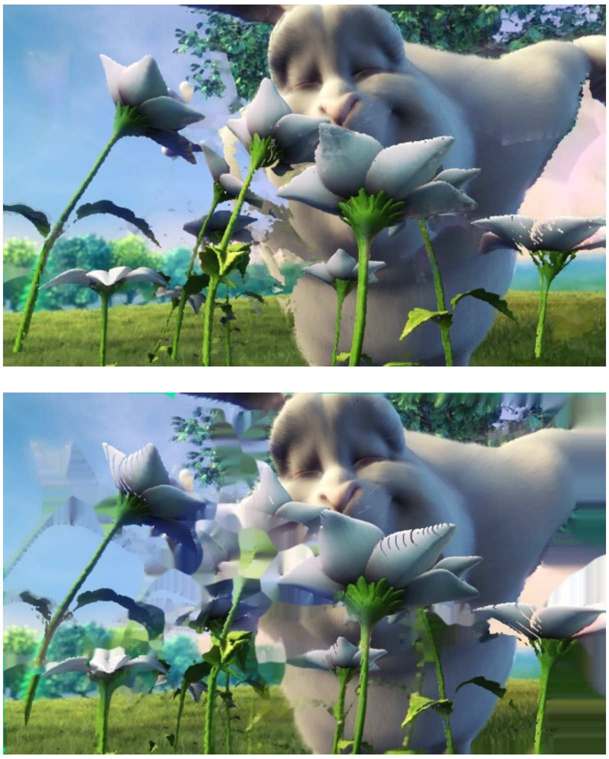


Fig. 10. Example resulting frames form the BBB Flowers sequence. Method described in (III.A.2) on the top, (III.B.1) on the bottom.

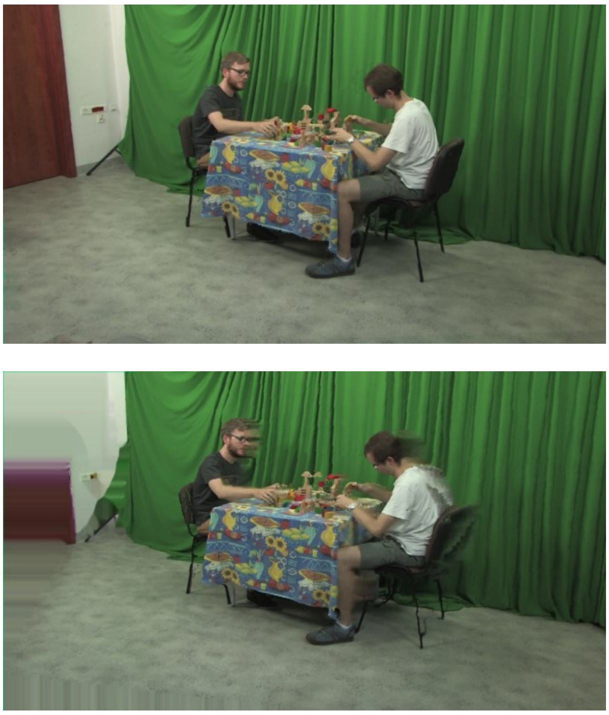


Fig. 11. Example resulting frames form the Poznan Block sequence. Method described in (III.A.2) on the top, (III.B.1) on the bottom.

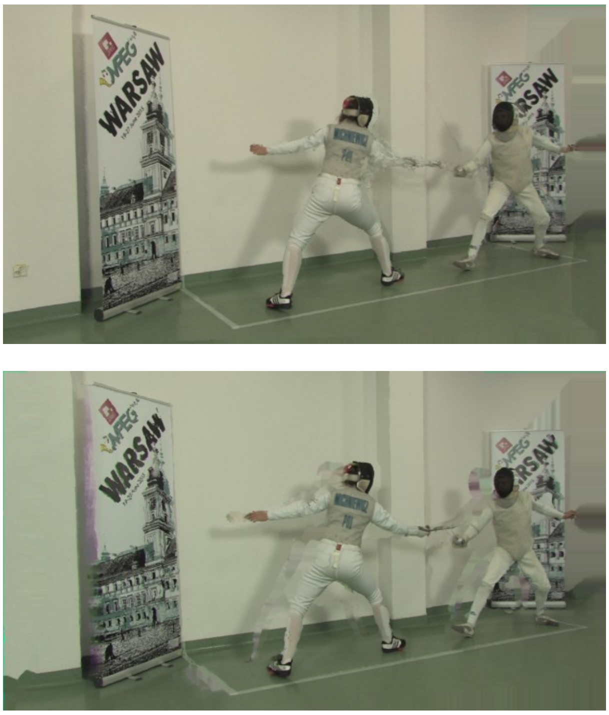


Fig. 12. Example resulting frames form the Poznan Fencing sequence. Method described in (III.A.2) on the top, (III.B.1) on the bottom.

VI. Summary

In this paper, four approaches to creating a high frame-rate video from a low frame-rate video, acquired by a multicamera system, have been analyzed. The paper shows that is possible to employ a regular multicamera system to create a high frame-rate video by utilizing virtual view synthesis and video sequences captured with a standard frame rate by multicamera systems.

VII. Acknowledgments

Project funded by The National Centre for Research and Development in the LIDER Programme (LIDER/34/0177/L-8/16/NCBR/2017).

VIII. References

[1] Dehghannasiri, R., Reza Soroushmehr, S.M., Shirani, S., “Frame rate up-conversion using nonparametric estimator,” 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 2014, pp. 3872-3876.

[2] Domański, M., Dziembowski, A., Grzelka, A., Mieloch, D., Stankiewicz, O., Wegner, K., “Multiview test video sequences for free navigation exploration obtained using pairs of cameras,” ISO/IEC JTC1/SC29/WG11, Doc. MPEG M38247, Geneva, 2016.

[3] Domański, M., Dziembowski, A., Mieloch, D., Wegner, K., Stankiewicz, O., “Integration of multiple input views into the View Synthesis Reference Software,” ISO/IEC JTC1/SC29/WG11 MPEG2018, doc. M42941, Ljubljana, Slovenia, July 2018.

[4] Dziembowski, A., Grzelka, A., Mieloch, D., Stankiewicz, O., Wegner, K., Domański, M., “Multiview Synthesis – improved view synthesis for virtual navigation,” 32nd Picture Coding Symposium (PCS), Nuremberg, Germany, 2016.

[5] Dziembowski, A., Mieloch, D., Stankiewicz, O., Domański, M., Lee, G., Seo, J., “Virtual View Synthesis for 3DoF+ Video,” 2019 Picture Coding Symposium (PCS), Ningbo, China, 2019.

[6] Fehn, C., Barre, R., Pastoor, S., ‘‘Interactive 3-DTV-conceptsand key technologies,’’ Proc. IEEE, vol. 94, no. 3, pp. 524–538, Mar. 2006.

[7] Kang, Y.-S. Ho, Y.-S., “Disparity map generation for color image using TOF depth camera,” 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), Antalya, Turkey, 2011.

[8] Kovacs, P., “[FTV AHG] Big Buck Bunny light-field test sequences,” ISO/IEC JTC1/SC29/WG11, Doc. MPEG M35721, Geneva, 2015.

[9] Kurc, M., “Hybrid techniques of depth map estimation and their application in three-dimensional video systems,” PhD Dissertation at Poznan University of Technology, 2019.

[10] Mieloch, D., Grzelka, A., “Segmentation-based method of increasing the depth maps temporal consistency,” Int. Journal of Electronics and Telecommunications, vol.64, no.3, pp.293-298, Warsaw, Poland, 2018.

[11] Motion estimation and compensation plugin for Avisynth+ and Avisynth v2.6 family, http://www.avisynth.nl/users/fizick/mvtools/mvtools2.html, online access 2019.

[12] Park, K., Kim, S., Sohn, K., “High-precision depth estimation using uncalibrated LiDAR and stereo fusion,” IEEE Trans. Intelligent Transportation Systems, vol.21, no.1, pp.321-335, Jan 2020.

[13] Qin, Y., Jin, X., Chen, Y., Dai, Q., “Enhanced depth estimation for hand-held light field cameras,” IEEE Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP), New Orleans, LA, 2017.

[14] Song, Y., Ho, Y.-S., “Time-of-flight image enhancement for depth map generation,” Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, South Kore, 2016.

[15] Song, Y., Ho, Y.-S., “High-resolution depth map generator for 3D video applications using time-of-flight cameras,” IEEE Trans. Consumer Electronics, vol. 63, Nov. 2017.

[16] Stankiewicz, O., Domański, M., Dziembowski, A., Grzelka, A., Mieloch, D., Samelak, J., ‘‘A free-viewpoint television system for horizontal virtual navigation,’’ IEEE Trans. Multimedia, vol. 20, no. 8, pp. 2182–2195, Aug. 2018.

[17] Tanimoto, M., Tehrani, M. P., Fujii, T., Yendo, T., “FTV for 3-D spatial communication,” Proc. IEEE, vol. 100, no. 4, pp. 905-917, 2012.

[18] Tsai, T., Lin, H., "Hybrid Frame Rate Upconversion Method Based on Motion Vector Mapping," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 11, pp. 1901-1910, Nov. 2013.

[19] Wang, T.S., Choi, K.S., Jang, H.S., Morales, A.W., Ko, S.J., “Enhanced frame rate up-conversion method for UHD video,” IEEE Transactions on Consumer Electronics, vol. 56, no. 2, pp. 1108-1114, 2010.

[20] Wilburn, B., Joshi, N., Vaish, V., Levoy, M., Horowitz, M., “High-speed videography using a dense camera array,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, Washington, USA, 2004.