IEEE Conference on Telepresence (Telepresence), Leiden, Netherlands, 2025.

Foveated Compression for Immersive Telepresence Visualization

Max Schwarz* and Sven Behnke

***Abstract*— Immersive televisualization is important both for telepresence and teleoperation, but resolution and fidelity are often limited by communication bandwidth constraints. We propose a lightweight method for foveated compression of immersive televisualization video streams that can be easily integrated with common video codecs, reducing the required bandwidth if eye tracking data is available. Specifically, we show how to spatially adjust the Quantization Parameter of modern block-based video codecs in a adaptive way based on eye tracking information. The foveal region is transmitted with high fidelity while quality is reduced in the peripheral region, saving bandwidth. We integrate our method with the NimbRo avatar system, which won the ANA Avatar XPRIZE competition. Our experiments show that bandwidth can be reduced to a third without sacrificing immersion. We analyze transmission fidelity with qualitative examples and report quantitative results.**

I. Introduction

Telerobotics is an important field in robotics which receives increasing attention in recent years [1]–[7], fueled by the availability of high-quality sensors and displays, high-bandwidth network connections, and advances in haptics and audiovisual telepresence. Increasingly, teleoperation is used to teach robots complex capabilities, producing teacher signals for learning [8]–[10].

Telepresence systems must capture the remote scene and display it to the operator in an immersive way. To this end, the most important human sense to be transmitted is vision. Since the human visual system is capable of resolving very fine details, high resolution image data needs to be captured on the robot side and displayed to the remote operator. The communication network between robot and operator is often severely limited in bandwidth, though. For example, a search & rescue robot inside a building might have to rely on low-bandwidth wall-penetrating wireless technology.

However, the human eye has variable acuity, with most resolution dedicated to the central region, the fovea [11]. In the periphery, resolution is drastically reduced—meaning that any high-resolution image data displayed there serves no purpose. This insight is not new, especially in VR contexts, where foveated rendering [11] and foveated video compression and streaming [12]–[15] are common techniques to reduce bandwidth and/or computational load.

In this work, we propose to use foveated compression for televisualization in the context of robotic telepresence systems to reduce bandwidth demands without adverse effects on immersion. To this end, we use eye tracking information to predict the operator's gaze direction, project it into the camera image using kinematic information, compute a Quantization Parameter (QP) delta map and feed it into the HEVC video codec to increase lossy compression in the periphery.

*Both authors are with the Autonomous Intelligent Systems group of University of Bonn, Germany; the Lamarr Institute for Machine Learning and Artificial Intelligence; and the Center for Robotics at University of Bonn. Contact: schwarz@ais.uni-bonn.de

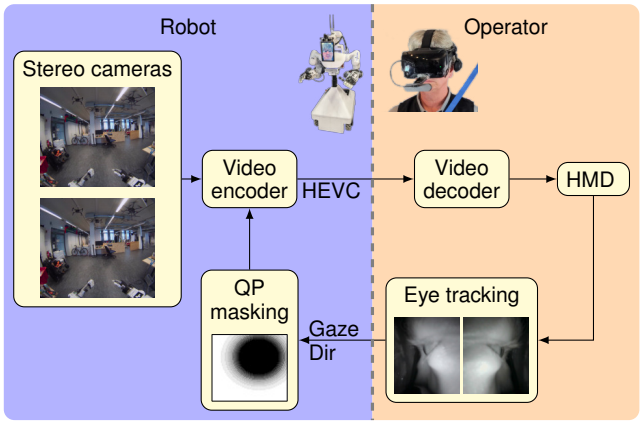


Fig. 1. Overview of our method. We provide immersive visualization of the robot's environment to the operator by streaming stereo images to the operator's HMD. For transmission over a potentially bandwidth-limited network, the video stream is encoded using HEVC. We propose to utilize eye tracking information to modulate HEVC's Quantization Parameter (QP) spatially—decreasing quality (and thus bandwidth usage) in areas that are not in the focus of the operator.

We integrate our method in the NimbRo avatar system [16], the system that won the ANA Avatar XPRIZE Challenge [17]. It features a movable 6-DoF head with a high-resolution stereo camera system.

We evaluate the integrated system with qualitative and quantitative experiments. Our results show that bandwidth can be reduced significantly without sacrificing immersion. We analyze transmission fidelity and report performance metrics.

In short, our contributions include:

1) A lightweight technique for foveated compression that can be easily integrated with common video codecs,
2) integration of the method with the NimbRo avatar system, and
3) qualitative and quantitative experiments demonstrating its effectiveness.

II. Related Work

Related works comprise saliency-based video compression, foveated rendering, and foveated streaming.

a) Saliency-Based Video Compression: A larger corpus of related work focuses on predicting human visual attention in order to focus compression quality on attended regions [18], [19]. Usually, no eye tracking is involved and the focus point is generated purely from predicted saliency

in the image. Lyudvichenko *et al.* [19] describe a hybrid method that combines eye-tracking data captured from a human observer with a model of spatiotemporal attention to predict multiple focus points. In contrast to our method, the above works focus on offline video compression and are thus not directly suitable for real-time usage.

b) Foveated Rendering: Foveated Rendering aims to reduce computational load for rendering VR scenes by reducing rendering quality in the peripheral areas. Wang *et al.* [11] provide an overview of the field. Similar to our method, foveated rendering utilizes eye tracking hardware built into the HMD to estimate gaze direction. While foveated rendering can be employed for VR telepresence, it can only reduce computational load on the operator side.

c) Foveated Streaming: In order to reduce network bandwidth, the quality reduction needs to happen earlier in the pipeline on the producer side (as in our work, see Fig. 1). Kämäräinen and Siekkinen [12] follow this idea, but for remote rendering, where VR scene generation is moved to the cloud instead of the client. Pohl *et al.* [13] also describe a foveated compression scheme for (2D) video streaming, where frames are divided into nine rectangular regions according to eye tracking information. The central region is transmitted with full resolution, while resolution is reduced for the periphery. In contrast, our method allows much more fine-grained spatial control of compression quality. Kaplanyan *et al.* [14] propose DeepFovea, a method for reconstructing video from a foveated encoding that retains only sparse color information in the periphery. In contrast to our method, which works with common video codecs, it is not immediately clear how to compress and transmit video efficiently with this representation. Bezugly *et al.* [15] propose a lightweight pre- and postprocessing stage that offers foveated compression around a video encoding. It works by warping the input image in such a way that the foveal region is retained in its original form, but the periphery is compressed into less space. The warped image is encoded and decoded as usual by the video codec and then de-warped. This approach controls bandwidth by varying resolution instead of compression levels. However, it suffers from warping artifacts such as aliasing effects. Lungaro *et al.* [20] implement foveated compression for video transmission by overlaying a low-resolution background stream with high-resolution foreground tiles selected by the user's gaze. In contrast, our method offers higher spatial resolution of quality control. Wu *et al.* [21] present gaze-driven perception-aware volumetric content delivery for mixed reality headsets employing a log-polar transformation around the fixation point for 2D content augmentation. Li *et al.* [22] introduce a log-rectilinear transformation to enable foveated streaming of 360° videos with off-the-shelf video codecs for VR headsets with eye-tracking. Illahi *et al.* [23] survey foveated rendering, encoding, and warping techniques for foveated streaming. In earlier work, Illahi *et al.* [24] propose a very similar QP modulation scheme as the one presented in this paper, but applied to video streaming for remote gaming. Tefera *et al.* [25] use foveated compression for RGB-D point clouds in a third-person robot telemanipulation setting. However, point clouds are difficult to display in dense fashion, especially when subsampled. Furthermore, RGB-D sensors often fail on reflective or transparent materials.

To our knowledge, there is no prior work on foveated video compression for immersive telepresence in robots.

III. Method

A. NimbRo Avatar Telepresence System

We base our work on our NimbRo Avatar system [16], which consists of an anthropomorphic robotic avatar and an accompanying operator station and strives for full audiovisual and haptic immersion. It was designed for the ANA Avatar XPRIZE competition [6], [17], where our team received the highest score in the semifinals and won the grand prize in the finals.

B. High-resolution Wide-angle Stereo Cameras

The avatar robot is equipped with a 6-DoF-movable head with two Basler a2A3840-45ucBAS cameras in stereo configuration. Fisheye lenses are mounted on the cameras to provide a field of view of more than $180°$. Images are captured at 47 Hz with a resolution of 2472×2178 pixels and processed in real-time in the onboard computer.

The images are captured and transferred in a native 12-bit Bayer format to the Nvidia RTX 3070 GPU, where they are processed using CUDA. After unpacking, automatic white balancing, and debayering, the RGB images are passed through a bilateral filter with $\sigma_d = 1.1$ and $\sigma_r = 0.01$ to reduce remaining debayering artifacts. After exposure and tone mapping (ACES Film), we perform color correction in the Oklab color space [26] and finally convert to the sRGB color space through gamma correction. The operations are implemented as CUDA kernels and fused as far as possible to guarantee minimum latency.

C. Video Codec & Network Transmission

The video frames are encoded using the Nvidia NVENC HEVC codec directly on the GPU. We utilize NVENC's `ULTRA_LOW_LATENCY` preset to achieve minimum latency and use intra refresh rather than IDR frames to guarantee constant bandwidth. The compressed frames are then transferred to the CPU and transmitted using the `nimbro_network` library which provides robust network transport over unreliable links for the ROS1 ecosystem[1]. For details on the NimbRo avatar system's WiFi transmission, we refer to Lenz *et al.* [16].

On the operator side, the compressed video frames are received and uploaded to the GPU of the operator PC, an Nvidia RTX A6000. The frames are decoded using the built-in NVDEC hardware decoder and directly passed to the VR rendering component.

D. HMD & Eye Tracking

In our work, we use a Valve Index HMD, offering $2\times1440\times1600$ resolution at up to 144 Hz with $108°$ horizontal field of view. In earlier work [27], we modified the

[1] `https://github.com/AIS-Bonn/nimbro_network`

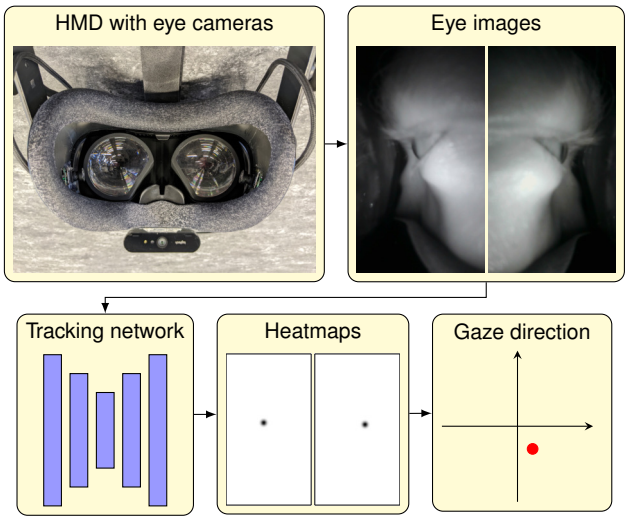


Fig. 2. Eye tracking pipeline. We mount additional eye cameras into an off-the-shelf HMD (Valve Index). The captured eye images are processed by a lightweight hourglass network to heatmaps describing an eye keypoint. Finally, the detected keypoints are transformed into the eye coordinate system and averaged to determine the gaze direction.

HMD to add eye cameras and a mouth camera to capture facial dynamics of the person wearing it.

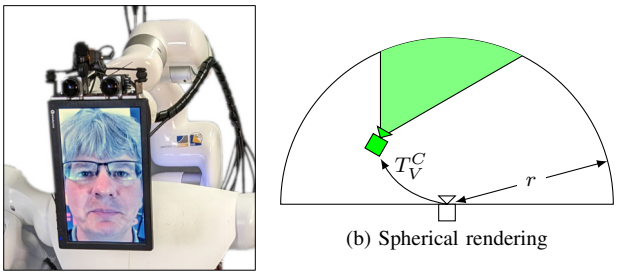
For clarity, we briefly explain the eye tracking component here (see Fig. 2). We mounted two miniature cameras inside the HMD that look from the sides inwards. Illumination is provided by IR LEDs without distracting the user. The system is initially calibrated by having the user follow a moving red dot with their eyes. We then train a lightweight hourglass network with only two downsampling and two upsampling layers that processes each eye camera image and outputs a heatmap describing the eye keypoint location. This keypoint is supervised only through a learned transformation from the camera coordinate system into the eye coordinate system, where calibration data is available. Calibration and network training can be done in roughly one minute.

During operation, predicted gaze directions from left and right eye are averaged to produce a single focus ray.

Our method does not depend on details of this eye tracking method. Therefore, off-the-shelf eye tracking systems can also be used.

E. 6-DoF Camera Movement & Spherical Rendering

A unique property of the NimbRo Avatar system is the movable 6-DoF head, which follows the operator's head motion with 1:1 correspondence. This feature generates proper parallax and allows looking around objects and easily moving to new viewpoints without repositioning the robot. However, there are latencies in the system from network transmission and the 6-DoF neck mechanism moving the avatar's head. This latency is compensated by rendering new views with low latency on the operator side under the assumption of constant distance (see Fig. 3). In essence, each video frame is rendered by projection onto a sphere centered onto the camera position at time of capture, allowing the VR camera to move freely in this sphere. This approach



(a) 6-DoF head with stereo cameras

(b) Spherical rendering

Fig. 3. Avatar robot head & Spherical rendering. (a) The NimbRo avatar robot head with a 6-DoF neck and stereo cameras. (b) Spherical rendering example. We show only one camera C of the stereo pair, the other is processed analogously. The robot camera is shown in white with its very wide FoV. The corresponding VR camera V, which renders the view displayed to the operator, is shown in green. The camera image is projected onto the sphere with radius r, and then back into the VR camera. Adapted from Schwarz and Behnke [28].

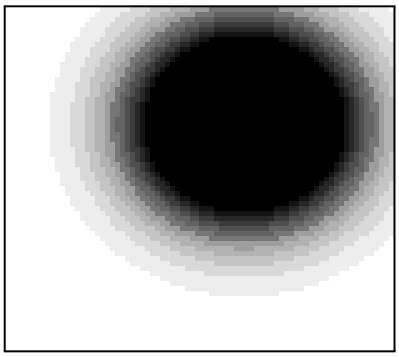


Fig. 4. Exemplary QP modulation mask for an image of size 2472×2178, resulting in a mask size of 78×69. White corresponds to a high QP change (yielding higher compression) while black indicates no QP change.

hides movement latency and is unnoticeable except for large head translations which reveal wrong distance assumptions. Note that it is beneficial to have substantially larger field of view in the camera system ($> 180°$) than in the HMD ($108°$) in this setup, since the operator can turn their head quickly and would see past the field of view in that case, breaking immersion. We refer to Schwarz and Behnke [28] for further details and a user study which shows the benefit of the movable head compared to standard pan/tilt mounting and fixed mounting.

F. Quantization Parameter Modulation

The Quantization Parameter (QP) of the HEVC codec directly determines the quantization of individual blocks and thus influences the compression factor. An increase of the QP by one corresponds to a 12% increase in the quantization step size, resulting in 12% decrease of the bitrate for the same content. NVENC allows specifying a *QP delta* Δq for each macroblock, which in our case corresponds to 32×32 pixels.

We choose to use a spatial Gaussian to produce Δq. While this does not closely match the behavior of human visual acuity [11], it allows us to easily define a wider foveal area that tolerates errors in eye tracking as well as latencies

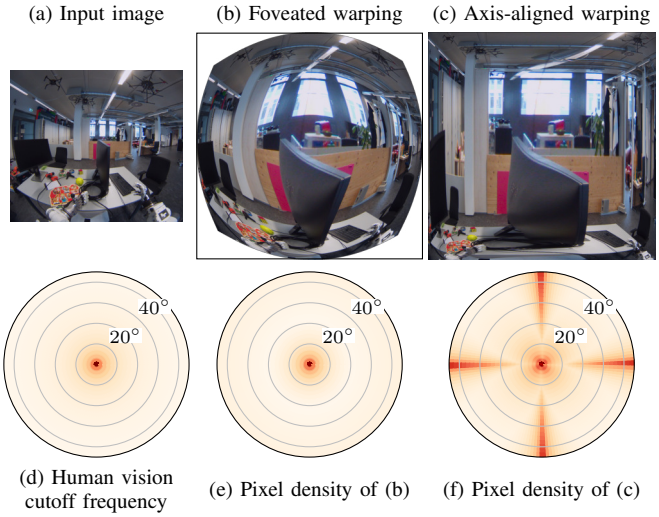


Fig. 5. Foveated warping example. The input image (a) is warped so that pixel density, visualized in a polar fixation-centric coordinate system (e) follows the cutoff frequency of the human vision system (d). However, this wastes space in the corners. By restricting the warping to individual axes, the entire space can be utilized (c), however this introduces additional modes in pixel density (f). In the lower plots, darker color indicates higher frequency or density.

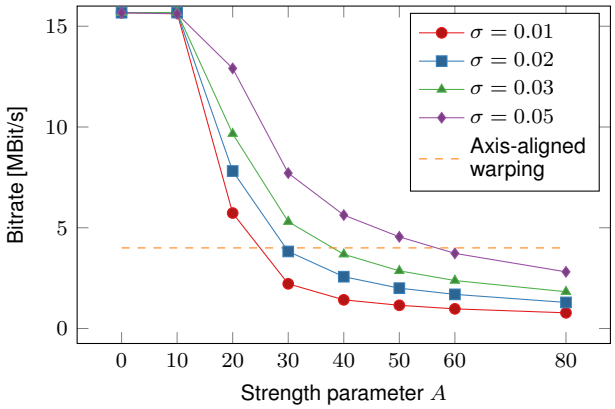


Fig. 6. Bandwidth reduction in foveated compression dependent on strength parameter A and spatial spread σ. For comparison, a reasonable bitrate for the axis-aligned warping method is shown.

introduced by network transmission and camera frame rate. We thus define

$$\Delta q(p) = A\left[1 - \exp\left(-\frac{||p-\mu||_2^2}{2\sigma}\right)\right], \tag{1}$$

where p is the normalized location in the image, μ is the projected gaze direction, and σ and A control spread and strength of the quality reduction, respectively. Fig. 4 shows an exemplary QP mask. We recommend $A = 50$ and $\sigma = 0.03$ (see Section IV-D).

G. Foveated Warping

For qualitative and quantitative comparisons, we also implement an alternative method for foveated bandwidth reduction called *Foveated Warping* [23], where images are transformed in such a way that areas close to the gaze fixation point are represented with more pixels, while peripheral areas receive less space in the image. Consequently, the total image resolution can be reduced, which will reduce video bandwidth automatically.

While having the advantage that the local compression effect can be modulated with higher spatial resolution (unlimited by the macroblock size), this approach also has drawbacks: The warping introduces additional sampling effects and requires quasi-random memory access. Furthermore, the resulting video stream is not compatible with standard viewers and requires de-warping—whereas the QP modulation method does not need any modification on the receiving side.

Nevertheless, we can apply foveated warping to our robotic telepresence scenario. In contrast to existing work, which mostly focuses on VR applications [23], our input images have fisheye characteristics, which have to be considered for warping.

As in related work [23], we use the model introduced by Geisler and Perry [29] for the cutoff frequency of the human vision system (HVS) depending on the angle θ (in degrees) to the fixation:

$$f(\theta) = \frac{e_2}{\theta + e_2}\frac{1}{\alpha}\ln\left(\frac{1}{\mathrm{CT}_0}\right), \tag{2}$$

where $e_2 = 2.3$, $\alpha = 0.106$, and $\mathrm{CT}_0 = \frac{1}{64}$ are the model parameters.

We then define the desired pixel density $g(\theta) = 4f(\theta)$, which ensures that the sampling theorem is satisfied (with margin for error). Finally, integrating over θ gives

$$G(\Theta) = \int_0^{\Theta} g(\theta)d\theta, \tag{3}$$

the required pixels for the angle range up to Θ.

For each output pixel (x, y) (centered around the fixation point) we can then compute $\theta_{(x,y)} = G^{-1}(||(x, y)||_2)$, the corresponding angle to the fixation point. Using $\theta_{(x,y)}$ the relevant 3D ray can be calculated, which is finally projected to the source image using the camera model.

A resulting warping can be seen in Fig. 5. To fully utilize the rectangular input shape of the video codec, the warping can be done for each axis individually. This deviates from the desired pixel density (which, however, is still a minimum bound).

IV. Experiments

We integrate our method in the NimbRo avatar system and report qualitative and quantitative experiments.

A. Bandwidth Reduction

Since the main motivation is bandwidth reduction, we measure bandwidth and show results in Fig. 6. We can see that bandwidth decreases linearly with rising strength parameter A at first, but there are diminishing returns—which is expected, since the foveal region needs to be transmitted at original quality and thus consumes constant bandwidth. A smaller σ also leads to small bandwidth savings, since less area is transmitted at high fidelity.

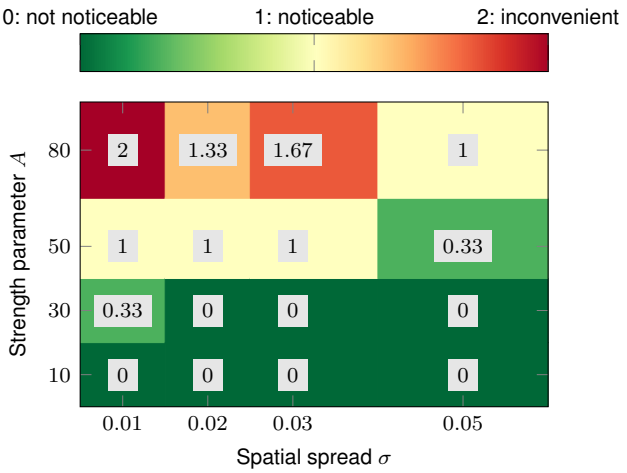


Fig. 7. User rating experiment. For each pair of (σ,A) the mean user rating is visualized. A rating of zero corresponds to "I did not notice any effect", one means "noticeable", and two means "inconvenient".

We note that the relationship between Δq and bitrate/bandwidth is complex, as initial QP values depend on scene complexity and content of each macroblock.

The comparison method (Foveated Warping) reduces the resolution from 2472×2178 (5.4 MP) to 1100×1100 (1.2 MP), greatly reducing the number of pixels. However, the warped regions contain high frequencies, which again increases the required video bitrate. We find that 4 MBit/s gives reasonable quality using the NVENC encoder.

B. Latency

We measure round-trip latency from an eye tracking measurement on the operator side to a video frame decoded on the operator side with the corresponding QP mask as 50 ms, which is in the same order of magnitude as saccadic omission [30], a neural mechanism inhibiting processing in the human visual system during rapid eye movements. Our visual system returns to its full capacity 60 ms after a saccade ends [31], giving time to update the display without noticeable latency.

The latency measurement was performed with Gigabit Ethernet connection between operator station and robot. Latency will increase with wireless connection ($<$ 10 ms additional latency) and with connection over the Internet (up to 300 ms), depending on network and distance.

C. Qualitative Examples

We show qualitative examples in Fig. 8. The effect of our method can clearly be seen as content near the gaze direction is preserved and quality is drastically reduced in the periphery. Compared to Foveated Warping, blocking artifacts can appear, although far away from the fixation point (see Fig. 9).

D. User Rating

To investigate which parameter settings of A and σ are acceptable for users, we performed an experiment with three inexperienced operators who where asked to rate specific settings with the following options: A rating of 0 means that there is no noticeable difference to the baseline (i.e. no foveation). Next, one means there is a noticeable difference, and two says that the effect is inconvenient or breaks immersion. As can be seen in Fig. 7, larger settings of σ allow higher settings of A, as expected.

Problematic scores for low σ mostly resulted from limited eye tracking precision. In these cases, the true operator focus point was too far away from the foveal area defined by eye tracking and σ. Eye tracking precision may be limited due to slight slippage of the HMD after calibration. We recommend setting $\sigma \geq 0.03$ to avoid these issues.

E. Limitations

Increasing the strength parameter A too much will result in high compression rates, which will of course lead to artifacts at some point. Since each macroblock is quantized separately, the macroblock grid will become visible at high QP offsets (see Fig. 9)—though this is usually not noticeable in peripheral vision.

The NimbRo avatar system uses periodic intra refresh instead of IDR frames, i.e. instead of transmitting full self-contained frames in periodic intervals, the intra refresh method sweeps across the image and sends self-contained data only for a subset of the macroblocks at a time. However, we noticed that at high QP settings this sweep becomes visible and may be noticed as movement in the peripheral vision. We note that IDR frames are not a viable alternative, since they either lead to highly unstable transmission bitrate or, if bitrate is kept constant, result in a noticeable flash of lower-quality content across the entire frame. It would be conceivable to allocate more bits to the macroblocks affected by intra-refresh, but this does not seem to be possible with Nvidia NVENC. In early experiments, patching the NVENC encoder to make the selection of macroblocks for intra-refresh more random than the original linear sweep across the frame makes this effect less noticable.

If very high compression settings are desired, explicitly blurring peripheral regions after decompression could also reduce block effects and noticeable intra refresh.

V. Conclusion

We demonstrated a lightweight method for foveated compression of VR televisualization video streams. We showed successful integration with the HEVC video codec in the NimbRo avatar system. Qualitative and quantitative results indicate that bandwidth can be reduced to a third without noticeable artifacts that would reduce immersion.

For even higher bandwidth savings, the proposed method could be combined with Foveated Warping, which retains better quality far away from the fixation point. Another promising direction of future research could be the prediction of the fixation point during a saccade to mitigate latency.

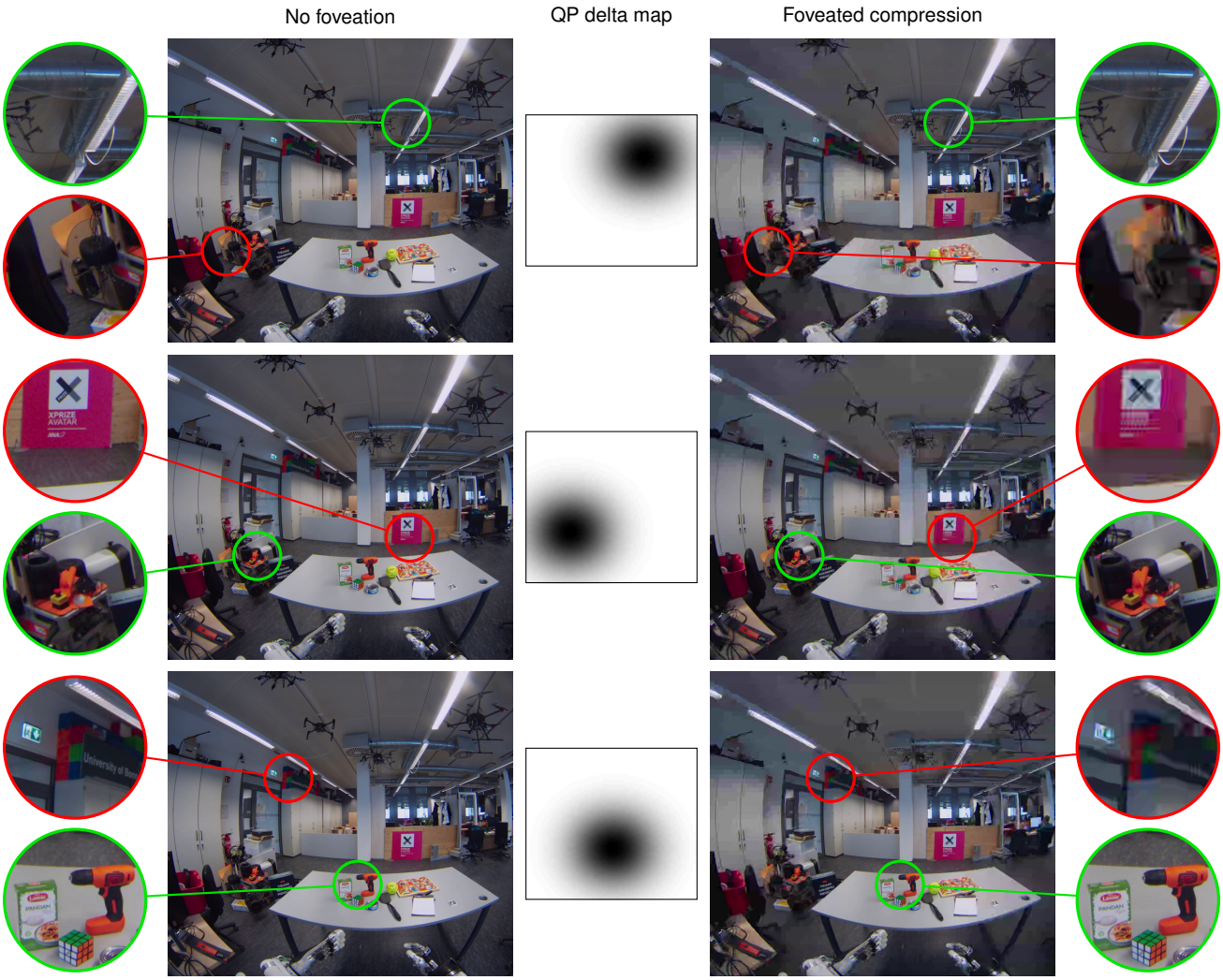


Fig. 8. Qualitative examples of Foveated Compression. We show enlarged views of points close to the gaze direction (green) and points in the peripheral area (red). All examples use the recommended values of $A = 50$, $\sigma = 0.03$.

ACKNOWLEDGMENT

This work was supported by the Robot Industry Core Technology Development Program (20023294, Development of shared autonomy control framework and AI-based application technology for enhancing tasks of hyper realistic telepresence robots in unstructured environment) funded by the Korean Ministry of Trade, Industry & Energy (MOTIE).

REFERENCES

[1] P. Stotko, S. Krumpen, M. Schwarz, *et al.*, "A VR system for immersive teleoperation and live exploration with a mobile robot," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2019.

[2] T. Zhou, Q. Zhu, and E. J. Du, "Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction," *Advanced Engineering Informatics*, vol. 46, 2020.

[3] T. Klamt, M. Schwarz, C. Lenz, *et al.*, "Remote mobile manipulation with the Centauro robot: Full-body telepresence and autonomous operator assistance," *J. Field Robotics (JFR)*, vol. 37, no. 5, 2020.

[4] M. E. Walker, T. Phung, T. Chakraborti, T. Williams, and D. Szafir, "Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 4, 2023.

[5] K. Darvish, L. Penco, J. Ramos, *et al.*, "Teleoperation of humanoid robots: A survey," *Trans. on Robotics (T-RO)*, vol. 39, no. 3, 2023.

[6] S. Behnke, J. A. Adams, and D. Locke, "The $10 million ANA avatar XPRIZE competition: How it advanced immersive telepresence systems," *Robotics and Automation Magazine (RAM)*, vol. 30, no. 4, 2023.

[7] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel, "GELLO: A general, low-cost, and intuitive teleoperation framework for robot manipulators," in *International Conference on Intelligent Robots and Systems (IROS)*, 2024.

[8] W. Si, N. Wang, and C. Yang, "A review on manipulation skill acquisition through teleoperation-based learning from demonstration," *Cognitive Computation and Systems*, vol. 3, no. 1, 2021.

[9] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning fine-grained bimanual manipulation with low-cost hardware," in *Robotics: Science and Systems (RSS)*, 2023.

[10] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile ALOHA: Learning bimanual mobile manipulation using low-cost whole-body teleoperation," in *Conference on Robot Learning (CoRL)*, PMLR, 2024.

[11] L. Wang, X. Shi, and Y. Liu, "Foveated rendering: A state-of-the-art survey," *Computational Visual Media*, vol. 9, no. 2, 2023.

[12] T. Kämäräinen and M. Siekkinen, "Foveated spatial compression for remote rendered virtual reality," in *First Workshop on Metaverse Systems and Applications (MetaSys)*, 2023.

[13] D. Pohl, D. Jungmann, B. Taudul, *et al.*, "The next generation of in-home streaming: Light fields, 5K, 10 GbE, and foveated compression," in *Federated Conference on Computer Science and Information Systems (FedCSIS)*, IEEE, 2017.

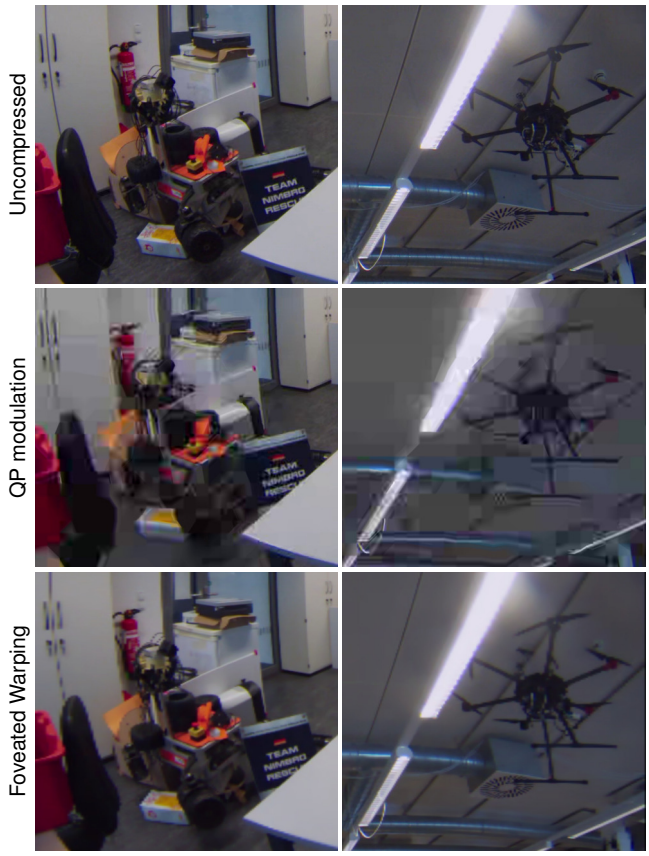


Fig. 9. Quality of QP modulation compared to Foveated Warping. Both crops are far away from the fixation point and demonstrate maximum quality loss.

[14] A. S. Kaplanyan, A. Sochenov, T. Leimkühler, M. Okunev, T. Goodall, and G. Rufo, "DeepFovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos," *Trans. on Graphics (TOG)*, vol. 38, no. 6, 2019.

[15] A. Bezugly, D. Rådell, and R. Biedert, "A lightweight foveation codec for VR," Tobii, Tech. Rep., 2021. [Online]. Available: `https://xr.io/static/docs/2021-12-01-Foveation.pdf`.

[16] C. Lenz, M. Schwarz, A. Rochow, *et al.*, "NimbRo wins ANA Avatar XPRIZE immersive telepresence competition: Human-centric evaluation and lessons learned," *International Journal of Social Robotics (SORO)*, vol. 17, no. 3, 2025.

[17] K. Hauser, E. Watson, J. Bae, *et al.*, "Analysis and perspectives on the ANA Avatar XPRIZE competition," *International Journal of Social Robotics (SORO)*, vol. 17, no. 3, 2025.

[18] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *Transactions on Image Processing (TIP)*, vol. 13, no. 10, 2004.

[19] V. Lyudvichenko, M. Erofeev, Y. Gitman, and D. Vatolin, "A semi-automatic saliency model and its application to video compression," in *International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2017.

[20] P. Lungaro, R. Sjöberg, A. J. F. Valero, A. Mittal, and K. Tollmar, "Gaze-aware streaming solutions for the next generation of mobile VR experiences," *Transactions on Visualization and Computer Graphics (TVCG)*, vol. 24, no. 4, 2018.

[21] N. Wu, K. Liu, R. Cheng, B. Han, and P. Zhou, "Theia: Gaze-driven and perception-aware volumetric content delivery for mixed reality headsets," in *International Conference on Mobile Systems, Applications and Services (MobiSys)*, ACM, 2024.

[22] D. Li, R. Du, A. Babu, C. D. Brumar, and A. Varshney, "A log-rectilinear transformation for foveated 360-degree video streaming," *Transactions on Visualization and Computer Graphics (TVCG)*, vol. 27, no. 5, 2021.

[23] G. K. Illahi, M. Siekkinen, T. Kämäräinen, and A. Ylä-Jääski, "Foveated streaming of real-time graphics," in *Multimedia Systems Conference*, 2021.

[24] G. K. Illahi, T. V. Gemert, M. Siekkinen, E. Masala, A. Oulasvirta, and A. Ylä-Jääski, "Cloud gaming with foveated video encoding," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 1, 2020.

[25] Y. T. Tefera, Y. Kim, S. Anastasi, P. Fiorini, D. G. Caldwell, and N. Deshpande, "Immersive remote telerobotics: Foveated unicasting and remote visualization for intuitive interaction," *Robotica*, vol. 42, no. 12, 2024.

[26] B. Ottosson, "A perceptual color space for image processing," Tech. Rep., 2020. [Online]. Available: `https://bottosson.github.io/posts/oklab/`.

[27] A. Rochow, M. Schwarz, M. Schreiber, and S. Behnke, "VR facial animation for immersive telepresence avatars," in *International Conference on Intelligent Robots and Systems (IROS)*, 2022.

[28] M. Schwarz and S. Behnke, "Low-latency immersive 6D televisualization with spherical rendering," in *International Conference on Humanoid Robots (Humanoids)*, 2021.

[29] W. S. Geisler and J. S. Perry, "Real-time foveated multiresolution system for low-bandwidth video communication," in *Human vision and electronic imaging III*, SPIE, vol. 3299, 1998.

[30] F. W. Campbell and R. H. Wurtz, "Saccadic omission: Why we do not see a grey-out during a saccadic eye movement," *Vision Research*, vol. 18, no. 10, 1978.

[31] L. C. Loschky and G. S. Wolverton, "How late can you update gaze-contingent multiresolutional displays without detection?" *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 3, no. 4, 2007.