DOI: 10.21062/mft.2023.057 http://www.journalmt.com

Devising a Multi-camera Motion Capture and Processing System for Production Plant Monitoring and Operator's Training in Virtual Reality

Joanna Gąbka (0000-0001-6550-1170)
Faculty of Mechanical Engineering, Wroclaw University of Science and Technology. Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland E-mail: joanna.gabka@pwr.edu.pl

The paper presents work aimed at building practical applications of virtual reality (VR) in manufacturing environments. It contains studies of the optical properties of cameras and lenses aimed at the selection of an optimal set (camera, adapter, lens) for the realization of recordings and video transmissions in stereoscopic format for VR. In response to the increasing trend in the number of applications of VR systems in the industry, works have been initiated with the purpose of building a system levelling image noise identified thus far as an obstacle to the effective utilization of VR in production systems. It was considered that picture error correction can significantly increase an already big data stream from the recordings. Based on it, a set of parameter values was defined which determined the selection of study equipment. Three research areas were set: the verification of the optical correctness, the study of image defects and their correction and the determination of the maximum optical resolution and the achievable image parameters in various lighting and environmental conditions. An example was presented for the application of a projected system for the monitoring of undesirable events/movement at work stands and key areas of production halls as well as training in the high risk production zones.

Keywords: Production monitoring, Optical properties measurement, Training Virtual Reality system in manufacturing, Image capture and processing

1 Introduction – relevance, areas of use, challenges

Virtual reality (VR) systems provide a new medium, which may have a significant impact on our society. It translates substantially into emerging trends in video systems, which start to be closely intertwined with the technologies of Virtual Reality and Augmented Reality (AR). This high tech is indicated in recent years as an increasingly important element of the Industry 4.0 development [1] especially in domain of training with 70% higher effectiveness in comparison to the traditional teaching methods. What is more, professionals gain skills using VR/AR systems, without being at risk of any immediate dangers[2]. VR/AR also supports other areas specified as crucial for Industry 4.0 such as building autonomus, robust and integrated production systems [3]. The main fields of modern manufacturing systems and engineering supported by VR/AR solutions are:

- Maitanance/Repair/Disassembly [4][5][6];
- Assembly [7][8][9];
- Design and Evaluation [10][11];
- Warehouse management [12][13];
- Quality control [14][15][16];
- Plant layout [17][18];
- CNC Simulation [19][20].

Each main pillar indicated [1][7] might be divided into sevral sub domains such as the quality control comprises tracking, monitoring and visualising the course of production process in real time [22] which might be also useful for production organising, control and management. Another intensly elaborated field is detailed parts quality inspection. The latter is especially important with wide range of applications from the quality of material surface evaluation to aiding process of sorting materials with VR system. The environmentally friendly solutions embedded in the production systems upgrade them towards Industry 5.0. The diversity of materials is mentioned as one of the biggest obstacles in recircular economy growth. The problem of material sorting technology is further described in [23] while the material surface quality challenges are raised in [24].

The experiences of VR/AR systems users differ considerably from the currently known reactions to television picture appearing on a screen [25]. The emergent solutions require efficient implementation of multi-camera setups to generate a high-quality image, an immersive 3D reality for the end user. The market sees the emergence of solutions supporting such applications as the monitoring of production processes, maintenance, the facilitation of quality monitoring systems, or the supervision over the security of facilities as well as training [26][27][28][29][30].

Under these circumstances, the development of solutions enabling the proper recording of stereoscopic images for virtual reality has become desirable element providing progress of their deployment in modern production systems and engineering. This area encompasses both the appropriate construction of the camera, algorithms for combining data streams from multiple cameras as well as the right formatting of displaying and display, which is optimized for capturing real scenes and events to ensure adequate user comfort and data quality enabling the use of algorithms i.a. for the recognition of movement or correct structure of item or material surface.

In order to select appropriate solutions, criteria have been developed that enable the assessment of the comfort of the user receiving the VR picture[31]. In particular, one should seek to meet the criteria defined as follows:

- Immersion – the viewer should feel immersed, i.e. present in the captured scene, the real world simulated by the VR technology ought to give a very strong sense of realness and genuineness, the users should be able to immerse themselves in the virtual world wearing 3D glasses and other tools for the effect of no distinction between reality and simulation as well as experience ongoing events [32], in a three-dimensional virtual environment built in the VR technology everything is supposed to look like real existence, what can be seen, heard, touched, smelled, and even tasted should be reflected in a real way, just like feelings in the real world;
- Real-time interactivity – interactivity concerns the fact that the user is located in a virtual environment, and the objects they touch and grasp are tactile, one can feel their weight, they can move along with the movement of human hands, which translates into high functionality of the whole environment;
- Stereopsis – in dynamic three-dimensional space, the user should feel any virtual spatial information according to their own perception and appraisal, they should have the possibility to use their own subjective perception enabling the acquisition of knowledge and information they need to cope and solve problems in space;
- Editing and streaming – the content should be represented in a form that can be edited with the existing tools, reliably streamed in today's networks and rendered in real-time in available setups.

To meet the defined criteria, a 3D image recorded based on a fisheye lens is used [33]. These solutions have been used in modern imaging products and have many industrial applications i.a. autonomous driving [34] or video surveillance [35]. The popularity of fisheye lenses stems from their large fields of view (FOV) up to 180° and more, which allows them to capture exceptionally large surrounding areas. They have considerable significance in creating static images in a VR environment as well as recording or emitting live images. This is a response to the emergent trend towards creating high-quality VR content with the use of VR 3D panoramic cameras. These cameras offer considerable benefits in terms of realism and immersion and are becoming more and more accessible thanks to numerous commercial possibilities. Also worth noting is live 360° video streaming, which will have enormous application in sport, theatre, telemedicine or video surveillance in industry based on VR sets. For VR to become well established in sport, theatre, telemedicine and telecommunication in general, live streaming is of key importance. At present, however, there is a significant difference in quality between live transmission and recorded VR content.

3D recordings are created with the use of multi-camera platforms, which require complicated and expensive optical sets necessary for the estimation of depth, the effective combination of multiple images and their rendering. Such sets may turn out to be too expensive in the case of a wide application in live transmission. As estimated in research, the capturing of content by the Google Jump VR camera requires 75-second processing per frame in the case of VR content [31]. Live transmission is therefore limited mainly to non-stereo 2D panoramic content. Recently, several 3D streaming cameras have appeared, but they require careful calibration, which may change in time, while their algorithms for real-time image combining may generate incorrect readings in the case of zooming in, reflecting, and transparent objects. Fig. 1 presents the process of data processing with the use of a camera set for VR image recording in the omni-directional stereo (ODS) format.

Camera sets for VR record high-resolution videos with multiple cameras, which requires considerable computing power and intensive processing. It includes the analysis of movement and particular views, which ultimately leads to the rejection of a significant amount of unprocessed data and their transformation into the omni-directional stereo (ODS) format.

There are several methods for recording VR video. An approach based on Light field rendering [36], which has been demonstrated with the use of two-dimensional camera matrices [37], showed that even if matrices such as these have been adapted for spherical capture, the amount of data that would have to be sent during live recording for the purpose of client-side rendering is huge.

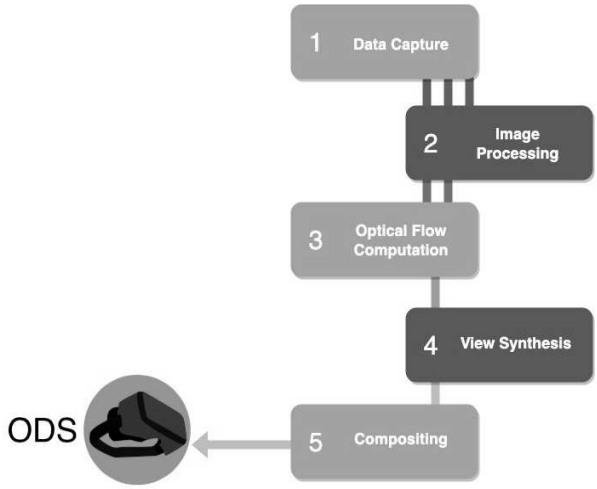


Fig. 1 *Model of data processing for systems based on camera sets*

An example of an effective image combination via sequence capture and smoothing is the ODS panorama. A challenge for ODS panoramas for live VR recording is the fact that stereoscopic depth cues ought to be operated in all possible viewing directions, which is not possible in the case of conventional panoramas. Capturing the light field offers guidelines concerning both the depth and the movement with 6 degrees of freedom, and recent works employing spherical lenses have demonstrated the light field capture with a wide field of view by a single lens [38]. However, the effective range of motion analysis offered by such a compact device is highly limited, which in effect requires multiple cameras to service any salient movement of a virtual camera.

A lot of effort has been put into the development of ODS capture systems for dynamic scenes due to the enhanced sense of immersion in the VR environment that can be achieved thanks to them. The system described by [39] is capable of obtaining video speed through high-speed rotating and capturing directly to the ODS format. However, the image quality of this system turned out to be low. It was only the complex system of mirrors and/or lenses implemented by [40] that enabled significant improvement in image quality.

Over the last several years, synchronized multi-camera systems have been advertised and offered by many companies in the consumer electronics industry. Systems comprising two spherical panoramic cameras [41] have been put forward, employing consumer electronics such as Ricoh Theta S or Samsung Gear 360 that is easily accessible but does not generate real omni-directional stereo images. Most ODS capture systems place cameras directly on the platform (16 in the case of Google Jump [31]) and use view interpolation to enable intermediate views between adjacent cameras. The greatest challenge for these systems is an enormous amount of captured data and impractical processing requirements. For example, Facebook's Surround 360 employs 17 high-resolution visual cameras recording with a speed of 30 frames per second. This system generates 17 Gb/s of unprocessed data, which are streamed via fibre optic connection to a large hard drive array with the purpose of storing and online processing. The most expensive processing stage is the low optical level between each pair of adjacent cameras and on many temporarily adjacent frames that enables smooth view interpolation. The authors [31] indicated that the necessary computing time equals 75 seconds for each VR video frame on a single computer with the use of their highly optimized algorithm or 75 computing days for a 1-hour video. Even with the use of clusters and hardware acceleration, live streaming is unattainable for these systems.

Among recent developments are also VR video cameras for live streaming, including Intel's True VR offering semicircular 3D video and Z-cam V1 Pro servicing 3D watching in 360 degrees with the use of NVIDIA's API5 real-time interface. These systems offer live streaming using accurate camera calibration as well as stereo-image stitching accelerated by GPU [42]. However, the efficiency of such systems may suffer with the presence of near, reflective and transparent objects, and the calibration may change in time and negatively affect the efficiency.

To overcome these limitations, Vortex architecture has been developed, which uses a mechanically moving construction that lacks sensitivity to manufacturing tolerance connected with exotic optical systems, requires neither computationally expensive interpolation nor stitching, is resistant to difficult scenarios and does not require significant calibration [43].

Stereo surround sound allows the user to look around but not move. It is consistent with a great part of existing VR glasses [44], which track the rotation of the head but do not track the translation. Correct handling of stereo sound is possible on the condition that the user does not turn her head (which most users do not do). Because what is emitted is a pair of panoramic films (one for the left eye and one for the right eye), the transmission requires only twice as much data as a monoscopic video, and many typical editing operations are easy (colour correction, dissolve etc.). There are also tools for creating more complex editions involving stereo [45]. For these reasons, the ODS format is used as a practical solution for providing VR video and the recording is performed based on cameras with fisheye lenses.

Literature also indicates challenges concerning the right presentation of VR images when watching with the use of VR glasses. These problems most often arise when the user performs different kinds of tasks while using VR glasses. The movement in distorted 3D virtual space may cause visually induced motion sickness [46]. Geometric distortions in stereoscopic 3D imaging may result from a mismatch between the parameters of capture, display, and watching of the image. Three pairs of potential mismatches are considered: 1) camera separation compared to eyes separation, 2) camera field of view (FOV) compared to screen field of view, and 3) cameras' convergence distance (i.e. the distance from the cameras to the point where the convergence axes intersect) compared to the distance between the screen and the viewer. What is also taken into account is the impact of the position of the user's head (i.e. the lateral displacement of the head from the middle of the screen). In response to the challenges related to the inadequacy of the video systems for VR designed so far, a study of five chosen hardware sets dedicated to the recording of video images for virtual reality has been conducted. Before devising a motion capture and processing system for virtual reality enabling a multi-channel transmission, a set of requirements was identified which must be met in order to ultimately obtain the possibility of real-time transmission and image correction [47]. Three research areas were set: the verification of the optical correctness, the study of image defects and their correction (deviations from the F-theta model, field of view angle of the lens, resolving power (MTF50) and lateral chromatic aberration), the determination of the maximum optical resolution and the achievable image parameters in various lighting and environmental conditions.

The literature review shows great potential of virtual reality solutions especially in the production plant monitoring on different levels and the training [26]. Both indicated areas respond to the common trend of decreasing number of stuff necessary for the process execution which results both from thriving to production plant autonomy and the lack of qualified workers available. Both of the challenges might be answered by the solution designed in this article. There are numerous studies showing that involving VR into the training process gives better results than conventional methods due to the immersion effect which enables operatrs to get awareness of challenges, dangers and general requirements occurring during the work process [48]. At the same time the studies showing benefits of VR training systems underline the importance of the wider view available and the quality of pictures available in realtime as a crucial factor in successful training processes and production plant monitoring both on macro level of the working station inspection as well as on the micro scale of assembly process commissioning. [49][50][51][52][53][54]. Beside the technical challenges tackled in this paper literature [55] names following obstacles: unavailability of experts, lack of sufficient cyber security solutions, financial barriers, insufficient awareness of VA/AR among stuff and lack of information and standardized processes of the VR systems implementation.

2 Studies of the optical properties of cameras and lenses towards effective VR system in manufacturing

In order to reflect the criteria described in the introduction, five parameters and their threshold values have been identified as key to the development of an effective multi-camera system generating stereoscopic video in semi-spherical format (180°) in real-time. They must be strictly met in order to effectively record and transmit a video stream in stereoscopic format. Systems used for displaying images in a selected format are commonly known as Virtual Reality (VR) systems. However, this term has a slightly broader sense since, apart from an image recording and projection set, it also includes equipment used for watching, especially VR glasses for the viewer. The article concentrates mainly on the elements used for recording, capturing, and transmission of a video stream. The study concerns sets consisting of three elements: camera, adapter, and lens. Their validation has been carried out based on a set of five key parameter values indicated below:

- minimum tonal range for the sensor in the camera set: 60db/10f– stops
- minimum resolution of the recorded image: 1000×1000 px
- minimum number of frames: 30 per second
- minimum bitrate for sending a stereoscopic image: 8000 kbps
- acceptable distortion of sets (camera, adapter, lens) when recording a stereoscopic semi-spherical image: +/- 20% from F – θ.

The study encompassed three areas:

- 1. The verification of the optical correctness of sets (camera, adapter, lens) when recording a stereoscopic 180° image.
- 2. The examination of occurring image defects and their correction (based on the methodology described in points 2.1.-2.4.

The measurement of the deviation from the F-theta model as well as the examination of the field of view angle of the lens were conducted. Moreover, the resolving power (MTF50) and the lateral chromatic aberration were verified.

- 3. The determination of the maximum optical resolution and the achievable image parameters in various lighting and environmental conditions. The examination of the dynamic range was conducted.

Five sets (camera, adapter, lens) were prepared to conduct the study:

- 1) Sony A7 III (ILCE-7M3) + Metabones MB-EF-E-BT5 adapter without optics + Canon EF 8-15mm f/4L Fisheye USM;
- 2) Panasonic Lumix DC-GH5S + Metabones Canon EF to Micro Four Thirds T Speed Booster XL 0.64z adapter with additional optics + Canon EF 8-15mm f/4L Fisheye USM;
- 3) Panasonic Lumix DC-GH5S + Fujinon FE185C086HA-1 with a passive adapter;
- 4) Pixelink PL-D755CU-BL-08051 + Sunex DSL-315B with a passive adapter;
- 5) Pixelink PL-D775CU-BL-08051 + Fujinon FE185C086HA-1.

Subsequent sections include descriptions of the assumptions made in conducting measurements and studies.

2.1 Deviation from the F-theta model. Examination of the field of view angle of the lens

An image recorded by means of cameras is often of insufficient quality to be directly transmitted and presented to the viewer. This is especially true when using cameras with wide field of view angles dedicated to obtaining stereoscopic (3D) images. In this case, a common problem is the so-called 'fisheye' effect – a distortion involving the constriction of visible elements along the edges of the lens. This effect is presented in Fig. 2.

F-theta lenses generate an image in an equidistant projection whereas it should be ultimately emitted in the more accessible rectilinear projection form (Fig. 3), when projected directly on a screen, or in the so-called equirectangular projection (Fig 4), popular for spherical and semi-spherical images (VR).

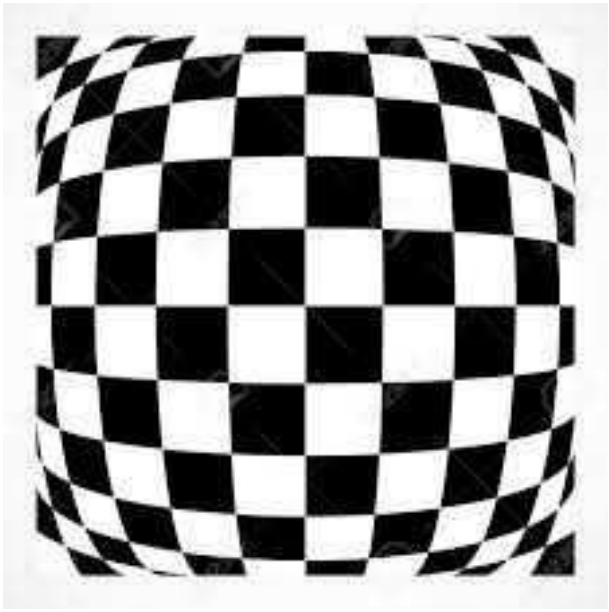


Fig. 2 *Radial distortion effect in wide-angle cameras*

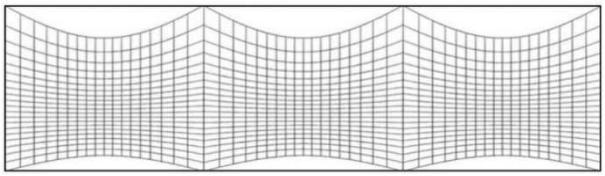


Fig. 3 *Grid of image reprojection to the rectilinear form, source: [56]*

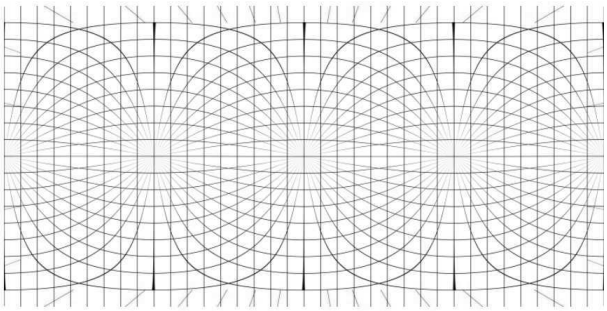


Fig. 4 *Grid of the equirectangular projection, source: [57]*

The reprojection of an image is understood here as the process of a change from one method of image display from a source to another. In this case, from equidistant into equirectangular and rectilinear. This operation should be performed before the transmission/projection of the image to the recipient. Owing to the high degree of complexity and the requirement for familiarity with the parameters of optics used in the recording, this cannot be assigned to the recipient of the image stream. In this connection, reprojection is a key element leading to the generation of a high-quality 3D image with an immersion effect for the viewer. There is no one universal method able to address the problem of distortion depicted in Fig. 2. [47] presents the overview and comparative analysis of the methods in terms of precision and computational efficiency connected with the necessity of reducing an

already considerable data stream. The described study uses a model, in which the transformation is implemented in three stages. Stage 1 consists in the calculation of the target projection. The result is the assignment of an entry vector V to each image pixel. Stage 2 encompasses the operation of rotating vector V and the reception of the resultant vector Vr. The last stage 3 constitutes the computation of the input projection where vector Vr is transformed into a pair of points lx, ly determining the point of the source image that is to be found in a given target pixel of the input image. For an image of the size of Pw×Ph, for each input pixel with coordinates Px, Py, a three-dimensional unit vector V is determined, specifying the direction from which an image is to be displayed at this point. It is also at this stage that the type of the target projection (equirectangular or rectilinear) is chosen.

Stage 1 for the equirectangular projection configured by the horizontal image area expressed as angle values in radians: $l \in [-\pi, 0]$, $r \in [0, \pi]$ and vertical $t \in \left[-\frac{\pi}{2}, 0\right], b \in [0, \frac{\pi}{2}]$. For an image of 180° vertically and 180° vertically, we receive values $l = -\frac{\pi}{2}, r = \frac{\pi}{2}, t = -\frac{\pi}{2}, b = \frac{\pi}{2}$.

In this case, the projection is first obtained through the calculation of the horizontal angle ϕ and the vertical angle θ. Where:

$$\varphi = l + P_x \frac{r-l}{P_w - 1} \, , \theta = t + P_y \frac{b-t}{P_h - 1} \tag{1}$$

Next, through the calculation of the unit vector V from them:

$$V = \begin{vmatrix} \sin\varphi \cos\theta \\ \sin\theta \\ \cos\varphi \cos\theta \end{vmatrix} \tag{2}$$

Whereas for the rectilinear projection configured by the viewing angle fr ∈ 0, π 2, vector V is obtained through the calculation of vector V ' representing the distance from the point on the projection plane where:

$$V' = \begin{vmatrix} P_x - \frac{P_w}{2} \\ P_y - \frac{P_h}{2} \\ \frac{w}{2 \times tg\frac{fr}{2}} \end{vmatrix} \tag{3}$$

Stage 2 The normalization of vector V:

$$V = \frac{V'}{|V|} \tag{4}$$

Given the values of vector V, it is possible to modify the camera's viewing direction by means of a simple rotation matrix R with the size 4x4.

$$Vr = R \times T \tag{5}$$

Stage 3 consists in the mapping of the equidistant (f-theta) projection of the input image. Each unit vector was assigned a point with coordinates Ix, Iy in the input image. Therefore the input image has a size of Iw×Ih. The configuration of the projection includes the viewing angle Fd ∈ [0, 2π], the displacement of the centre of the image Δx and Δy, aspect ratio a, and a chosen method of optics correction relative to the ideal f-theta lens (e.g. ABCD polynomial).

Parameter D introduced in the formulas indicates the direction of the lens. For calculation simplicity, it is permanently directed forward, hence its value can be expressed as follows [0,0,1]. If necessary, a result identical to the vector change D can be obtained by properly constructing matrix R from stage 2.

The distance from the centre of the lens r ' is calculated using the formula:

$$r' = D.Vr \tag{6}$$

The obtained distance r is then corrected by means of a polynomial or the data from the table. For the polynomial method with coordinates A, B, C, D, this is:

$$r = A \cdot r'3 + B \cdot r'2 + C \cdot r' + D \tag{7}$$

Apart from the distance, a two-dimensional direction vector d is determined, which is a normalized two-dimensional projection of vector Vr created through the deletion of the third coordinate and the normalization of the result by means of the formula:

$$d' = \begin{bmatrix} V_{r1} \\ V_{r2} \end{bmatrix} \; d = \frac{d'}{|d'|} \tag{8}$$

Given the corrected distance r, the normalized direction d (divided into components d1 and d2), the lens angle Fd and the image size Iw, Ih, input coordinates Ix, Iy are calculated:

$$I_x = \frac{d_1 \times r \times I_h \times a}{4 \times F_d} + \frac{I_w}{2} + \Delta x \; ; \; I_y = \frac{d_2 \times r \times I_h}{4 \times F_d} + \frac{I_h}{2} + \Delta y \tag{9}$$

The deviation from the F-theta model was measured by means of the Genesis Base MH-10 panoramic head attached to a stable tripod and enabling two-axis rotation. The head's tilting angle was set to 0° before adjoining the camera set in order to later obtain the parallelism between the optical axis of the lens and the ground. The head enables the rotation of the pan angle in the range of 360° with graduation of 2.5°. In order to eliminate the influence of the parallax on the measurement, the camera set

was collimated in such a way that the middle of the entrance pupil of the lens (point without parallax) coincided with the rotation axis of the head's panning. The camera set was directed towards a specially drawn point being a constant reference. Each measurement series was performed after the prior setting of the pan angle at 0° scale mark while ensuring that the reference point was in the centre of the frame of the recorded image. Each measurement was preceded by a tripod stability control. In each of the examined camera sets, the built-in corrections of the lens geometry, vignetting, and chromatic aberration were disabled (if they were available).

Measurement data have the form of photographs taken for various pan angles. In the initial range, where the distance of the recorded point from the optical axis is minor, the measurement resolution of 5° was adopted, while with the point located closer to the edge of the field of view (where one can expect the greatest deviations from the F-theta), the measurements were performed using the highest available head resolution of 2.5°.

Data analysis, after their collection, was performed on a computer through the comparison of images taken at different angular distances. With a given position of a point in the image for the 0° as well as its position after a rotation by a given angle, the distance in pixels of the point relative to the first measurement was calculated for each shift. The value of the difference between the measured distances relative to the rotation angle demonstrates the degree of the departure from linearity of a lens transformation. Aggregated measurement results are presented in Fig. 8-15 and in Tab. 5. In tables with partial results, part of the data in pixels are fractional values. This is due to the use of software that enables measurement with subpixel precision. In the case of a standard F-theta lens, a given angular shift would change the distance by the same number of pixels, regardless of whether the measured point is located in the centre or periphery of the frame. In real optical systems, this relationship deviates from a linear one, causing the radial distortion effect, which is expressed by means of a percentage deviation from F-theta. It was assumed that the field of view reference point is the maximum angle, for which the reference point was visible in the frame. The following possible sources of the main measurement errors that may affect the study were defined: the inaccuracy of reading and pan angle shift, the uncertainty of the setting of the camera set (parallax error).

2.2 Resolving power (MTF50)

Resolving power is described by means of the modulation transfer function (MTF). It specifies the response of the optical system depending on the spatial frequency. In practice, MTF is determined by comparing the level of contrast for low (where the signal is strong) and high frequencies (these refer to edges in an image). The spatial frequency is expressed in lines (or line pairs) per unit length, which can be a millimetre or a number of pixels. The value of the function corresponds to the number of details discernible by the optical system. In tests of camera sets, the unit used was Cy/mm (equivalent to lp/mm – lines per millimetre), meaning the number of line pairs that are distinguishable with adequate quality on a matrix millimetre. A derived unit also included in measurement results is LW/PH (line widths per picture height), which specifies the maximum number of lines possible to be reproduced on the vertical axis of an image, that is the height of the matrix. The commonly used measure of the perceived image sharpness is frequency, for which the contrast relative to the value drops by half for low frequencies. It is labelled as MTF50.

The low-frequency contrast is defined as:

$$C(0) = \frac{V_w - V_b}{V_w + V_b} \tag{10}$$

Contrast for the measured frequency f:

$$C(F) = \frac{V_{max} - V_{min}}{V_{max} + V_{min}} \tag{11}$$

The MTF50 function is represented by the formula:

$$MTF50(f) = 50\% \times \frac{C(f)}{C(0)} \tag{12}$$

Where: Vw – luminance of pixel value in the centre of the white line, Vb – luminance of pixel value in the centre of the black line, Vmax – the value of the brightest pixel for a given frequency, Vmin – the value of the darkest pixel for a given frequency.

The modulation transfer function specifies the aggregate resolving power of the components of a camera set – lens, optional adapter, and matrix. In theory, this enables the comparison of different sets. If single elements are compared, e.g. two different lenses, the remaining set components (e.g. the camera) did not differ. The lenses and cameras examined in this paper represent products designed for a variety of applications. Examples include the Canon 8- 15 camera lens drawing a full image circle on a Full Frame matrix of the area of ≈ 850 mm2 or the Sunex DSL-315 lens dedicated to closed-circuit television and designed for 85 mm2, hence approximately 10 times smaller, matrices. Due to this diversity and the application of additional optical elements (Metabones x 0.64), the analysis of the total optical possibilities of sets and their impact on video quality required taking into account the following set of factors:

- the physical size of the matrix, its resolution, and the physical size of the pixel,
- the optional use of an adapter with an additional optical element,

- the size of the image circle drawn and the actual use of the matrix area, the impact on the actual parameter LW/PH,
- the impact of the aperture size that differs between particular lenses.

Tab. 1 *An exemplary data set representing the time to failure (ti) of a given operational element*

Camera	Physical size of sensor [mm]	Sensor area [mm2]	Pixel size [µm]	Max. resolution [px]
Sony A7III	35.6×23.8	847.28	5.91	6000×4000
Panasonic GH5S	17.3×13	224.9	4.68	4096×2160
Pixelink PLD755CU-BL	8.8×6.6	58.08	3.45	2448×2048

Apart from the lens, a limiting factor to the resolving power of a video recording system can be the matrix. Its capability is specified by the Nyquist limit, i.e. the greatest theoretical frequency able to be reproduced by the sensor, equal to half its sampling rate. In practice, owing to non-zero distances between pixels and the possibility of beat frequencies, the Kell factor, set at 0.9 for CCD matrices, is used.

$$Nyquist\ limit = \frac{1}{2}\ x\ Kell\ factor\ x\ sampling\ rate \tag{13}$$

Theoretical Limits for each camera are presented in Tab. 2.

Tab. 2 *Nyquist limits for particular cameras*

	Nyquist limit [lp / mm]	Nyquist limit [LW/PH]
Sony A7III	76	3618
Panasonic GH5s	95.7	2488
Pixelink	125.18	1652

The procedure of measurement data acquisition was based on the taking of photographs of a test pattern in accordance with the ISO 12233:2000 standard, in A2 format (Fig. 5). The pattern was illuminated in a constant way with the use of continuous light photography lamps. The camera set was placed on a stable tripod, in such a way that the optical axis was perpendicular to the pattern while intersecting with its central point. In order to avoid vibration, each camera was triggered remotely. If the lens facilitated the change of aperture and/or focal length, measurement was performed for each possible setting.

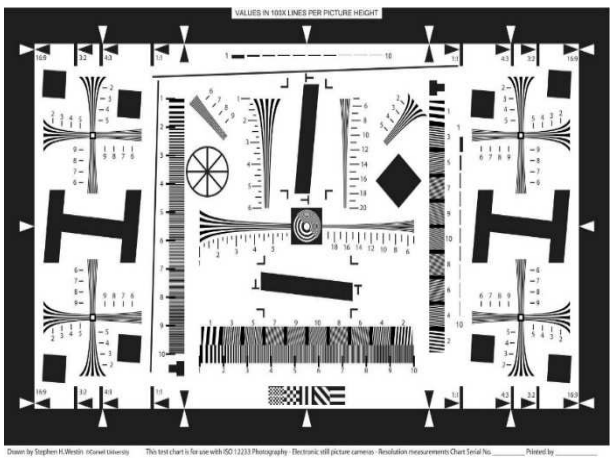


Fig. 5 *ISO 12233:2000 calibration board*

In studies of optical resolution, it is optimal to use RAW or BAYER RAW raw files (before the demosaicing process). Before the measurement procedure, they were converted to a 48-bit .tiff file. For the A7III camera, the dcraw program was used as recommended by the manufacturer of the Imates program. In the case of the GH5S (the dcraw program did not read the file properly), the darktable and Adobe Camera Raw programs were used (white balance set by default). Such settings as sharpening, chromatic aberration correction, distortion correction, and colour modification were disabled. For the Pixelink camera, .tiff files obtained by means of the manufacturer's software or a raw matrix screenshot (.bin format) were used. .jpg format files are additionally sharpened with an internal camera processor, which eliminates their usefulness for testing.

The analysis of the data obtained in this part of the study was conducted using the SFR (Spatial Frequency Response) algorithm implemented in the Imatest program. The presented results constitute an average of the MTF50 function in vertical and horizontal axes. Their calculation drew on inclined edges located near the centre of the test pattern's frame. The applied method enables the measurement of the values exceeding the Nyquist limit of the matrix thanks to 4x oversampling.

2.3 Lateral chromatic aberration

Chromatic aberration is one of the defects of optical systems necessary to be corrected for the fulfilment of the criteria of good-quality transmission and the immersion effect indispensable for VR systems. It involves the shift of the focusing of different light wavelengths (colours) relative to each other, which may cause the visibility of a rainbow rim at the edges. This phenomenon is enhanced near the boundaries of the field of view and is caused by dispersion, i.e. the dependence of the refractive index of a medium on the light wavelength. The minimization of the impact of the lateral aberration is one of the fundamental issues when designing an optical system. There are various methods of resolving this problem, which translates into different results obtained for the patterns of particular apertures and focal lengths. The visibility of chromatic aberration is proportional to the area between the edge with the highest amplitude (in this case the red curve) and the one with the lowest amplitude (the blue curve).

During the analysis of images coming from digital matrices, the natural measure of area expressing the chromatic aberration is the number of pixels. Such an approach assumes that the results are dependent on the resolution of the used camera and the place in the frame where the measurement was taken. Taking into account the fact that in most lenses the chromatic aberration is proportional to the distance from the lens centre, the measures were written in the form of the percentage value of the distance separating the middle of the image from the examined edge. For example, in the image (Fig. 6), the aberration CA(area) equalled 0.604 pixel value, which corresponds to 0.159% of the distance from the image centre to the measured edge. Such an approach enables the comparison of results for many cameras of different matrices as well as minimizes the impact of the different length of the examined edge from the frame centre. The chromatic aberration was measured using the same measurement photographs that were used for the resolving power tests. The slanted edge located at the edge of the ISO 12233:2000 calibration board was investigated. The measurements concerned camera sets characterized by different diameters of the view circle on the matrix, which is why, during positioning, efforts were made to ensure that the examined edge was at a similar distance from the frame centre for each circle. The manufacturer of the Imatest3 software shared data enabling a descriptive evaluation of the obtained aberration levels. They are included in Tab. 3.

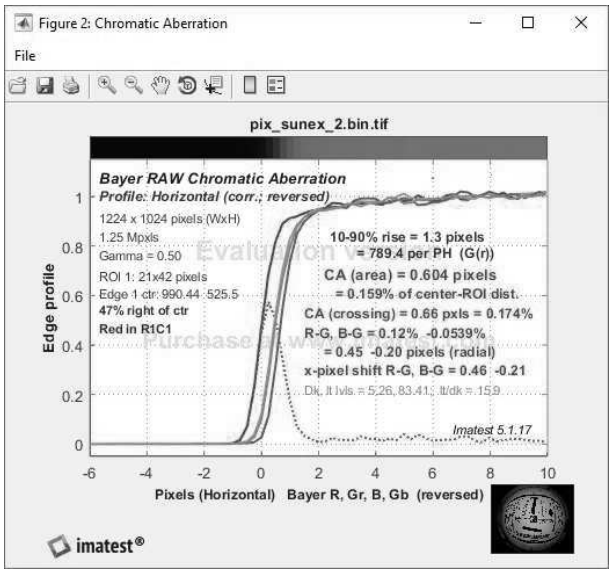


Fig. 6 *The process of the chromatic aberration at the edge; Graph generated in the Imatest program*

Tab. 3 *Perceived level of chromatic aberration*

Chromatic aberration as percentage distance from the middle of the frame	Descriptive level
0-0.04	Negligible
0.04-0.08	Low. Difficult to notice if not looked for
0.08-0.15	Moderate. Visible only on big enlargements
over 0.15	Considerable. Very visible on big enlargements

Chromatic aberration is a defect that arises only due to the imperfection of the optical system; however, it is the matrix and camera electronics that are responsible for its subsequent recording and processing. It generates the possibility of obtaining different results for the same lens working with different cameras, which is primarily dependent on the size and spatial distribution of pixels, and also the camera's processor (debayer algorithm). RAW photographs coming directly from the camera matrix were used for the chromatic aberration tests in order to bypass any image modifications made by the processor. The parameter's value is also influenced by the demosaicing process.

2.4 Tonal range

The tonal (dynamic) range is defined as the difference between the brightest and the darkest area of the captured frame, for which there is no detail (information) loss, meaning that the image is neither overexposed nor underexposed. The greatest the tonal range, the more contrasting shots can be recorded.

This relationship makes it one of the key parameters of a camera matrix. It is most often expressed by means of EV (Exposure Value) aperture units, where a one-unit increase means a two-fold increase in the intensity of the light incident on the matrix. A common unit is also decibel dB. The tonal range in chosen sets was measured by means of the Danes–Picta normalized reproductive processes quality control standard Q14. The imprinted grayscale has 20 grades, each of which differs in intensity by 1/3EV relative to its predecessor.

Fig. 7 *Normalized Danes-Picta Q14 Gray Scale standard for the measurement of tonal range*

The standard presented in Fig. 7 was evenly illuminated with continuous light studio lamps positioned in a way that eliminates reflection. Next, using each of the test cameras, a series of photographs of various expositions was taken in such a way that the brightest of them was overexposed and the darkest one clearly underexposed. When taking the photographs, the exposure time was manipulated, which is directly related to EV and characterized by precision and repeatability in establishing the value. The test pattern lay minimally beyond the focus plane in order to reduce the impact of noise coming from the pattern's texture. The ISO in the camera was set to a negative value. In order to obtain the most reliable results, RAW files were recorded with the greatest possible bit depth for each of the cameras. The gathered files with the results were converted into a TIFF file, where each channel was saved in 16 bits. The images were analyzed with the use of the 'Stepchart' module of the Imatest program. The determination of the tonal range was based on the quality criterion, where the signal noise ratio (SNR) related to the scene is greater than the minimum value related to the image. The greater the SNR (lower noise level) in a given region, the better the image quality. It was assumed that particular SNR values correspond to the quality in a way presented in Tab. 4.

Tab. 4 *SNR in the context of image quality*

SNR = 10	high quality (20dB)
SNR = 4	medium-high quality (12dB)
SNR = 2	medium quality (6dB)
SNR = 1	low quality (0dB)

3 Measurement results and comparative analysis

The first area indicated in the study outline was the verification of the optical correctness of sets (camera, adapter, lens) when applied in the recording of a stereoscopic 180° image. None of the examined sets had measurable backlash or changed its optical parameters upon displacement (horizontally, vertically) and the setting of the lens parameters (does not apply to set 4). All of them enabled proper focusing. In the first and second sets, active adapters were used, which effectively mediated between the lens and the camera, explaining commands and lens parameters, i.e. the parameters of aperture, focus, and the workings of the focus mechanism. In the remaining sets, the lenses do not have electronic elements. In the Fujinon lens, the focus is set with an Allen key, while in the Sunex one, with a nut. Additionally, Fujinon enables the selection of the focal length by means of a ring integrated with the lens. In set 4 (Pixelink + Sunex), discrepancies in positioning and image circle size between particular pieces of cameras and lenses were observed. Due to small image circle sizes (around 8 mm), the differences are negligible (as in sets 1 and 2). All discrepancies can be corrected at the stage of image processing through the calibration of the camera-lens pair with the values of the frame centre and the image circle size. None of the noted deviations reduced the image resolution or the lens angle of view below the data values listed in the section 2. The measurement results are following.

Canon 8-15 Lens is characterized by a high resolving power, whose maximum is achieved at an aperture of F/4. The choice of the aperture has a marginal impact on the chromatic aberration, which is negligible. In order to record a 180◦ image, an 8 mm focal length should be used.

Fujinon Lens the results of the test conducted within the second area of the study showed a high chromatic aberration for values F/1.8 and F/2. The resolving power did not yet achieve its maximum in this range, either. If lighting conditions permit, it is advisable to use apertures of F/2.8 and larger.

Sunex Lens does not have variable focal length or aperture. The correction of the middle of the image circle and its size is at the processing stage.

Sony Alpha 7 III Camera in HDMI preview mode (image not recorded to local data carrier) it is recommended to set the resolution to 4096×2160 pixel. Taking into account the image circle size that covers the whole length of the matrix height, the optimal resolution for further processing is 2160×2160 pixel, which retains the 1:1 ratio.

Panasonic Lumix GH5S Camera it is advisable to set the 4096×2160 pixel resolution and 59.94 frames per second. An anamorphic mode with an aspect ratio of 4:3, a resolution of 3328×2496, and 59.94 frames per second is also available.

Pixelink Camera enables free choice of resolution and frame rate. The maximum available resolution is 2448×2048 pixel with 60.8 frames per second. When

working with the Sunex lens (Set 4) on image capture, the correction of the possible image circle centre and the adjustment of resolution should be taken into account. For the Fujinon lens (Set 5), the recommended resolution while taking into account the image circle is 1900×1900 px.

The second study area was the analysis of image defects and their correction. In this area, the measurement of deviation from the F-theta model and the examination of the field of view angle of the lens were conducted. Moreover, the resolving power (MTF50) and the lateral chromatic aberration were verified.

The deviation from the F-theta model was measured for each camera set. The maximum deviation did not exceed 18% in any of them. A summary of the measurement results obtained is presented in Fig. 8.

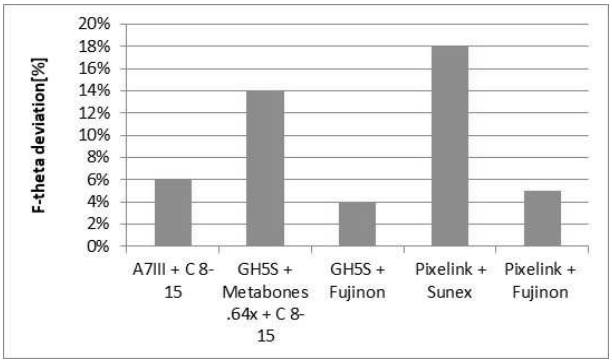


Fig. 8 *Comparison of the F-theta parameter for camera sets*

The difference in the obtained results between sets I and II (both involve the Canon 8-15 lens) is caused by the impact of the adapter with an additional optical element, which introduced additional distortion. The Fujinon lens was tested twice, in combination with the GH5S and Pixelink cameras. The difference between the results was equal to 1%, which can be considered a measurement error.

The measurement of the resolving power MTF50 was carried out for all five camera sets in the lp/mm and LW/PH units. The comparison of the resolving power of such different lenses and cameras working with them requires additional analysis. The value of MTF50 expressed in lp/mm is a measure of the number of lines possible to be drawn per matrix millimetre. Its use, as a measure of the quality of a camera set, is justified in the case of tests involving systems with sensors of identical size and comparable coverage. Using the obtained data, further analysis encompassing matrix sizes and priorly calculated percentage image circle coverage was conducted. A decision was made to abandon the classical definition of resolving power as the number of lines per unit length, replacing it with a dimensionless quantity specifying the capacity for the reproduction of details on the entire matrix surface. The obtained result multiplied by the percentage coverage enables the comparison of sets. Further calculations were performed with the simplification that the percentage decrease in the value of MTF together with the increase in the distance from the frame centre in particular sets is on the same level. Knowing the obtained results of LW/PH (the maximum number of lines possible to be reproduced on the image's vertical axis) and each camera's matrix size (Tab. 5), LW/PV was calculated (the maximum number of lines possible to be reproduced on the image horizontal axis).

The capacity for the reproduction of details (a dimensionless unit) was calculated using the formula:
MTF50max[LW / PH] × MTF50max[LW / PV] × Percentage Coverage

The results are summarized in Tab. 5, the value for A7III + C8-15 was taken as 100%.

Tab. 5 *Capacity for detail reproduction of particular sets*

Set	Maximum [LW/PH]	Maximum [LW/PV]	Coverage [%]	Capacity for detail reproduction	Capacity for detail reproduction % relative to A7III + C8-15
A7III + C 8-15 (8 mm)	2458	3664.7	50%	4.51E + 06	100%
GH5s +C 8-15 (8 mm)	1522	2025.4	81%	2.49E + 06	55%
GH5S + Fujinon	1979	2633	28%	1.45E + 06	32%
Pixelink + Sunex	1075	1433.3	71%	1.09E + 06	24%
Pixelink + Fujinon	968	1290.6	91%	1.14E + 06	25%

Fig. 9 presents the summary of the amount of detail possible to be reproduced by particular sets. The obtained results point clearly to the A7III + C8-15 set as capable of recoding the greatest amount of detail (information) in an image. The GH5s + C8-15 set reached the value of 55% compared to the first one. It is a very satisfying value taking into account an almost four times smaller matrix size of the GH5S camera. What contributed to this was greater coverage (81%) than in the first set as well as the enhancement of resolving power per millimetre. Both these values were obtained thanks to the use of the Metabones ×0.64 adapter. It adjusted the image circle size to the GH5S matrix and increased the MTF in line with the manufacturer's declarations. The comparison of the first and second sets in terms of the MTF50 value is presented in Fig. 10.

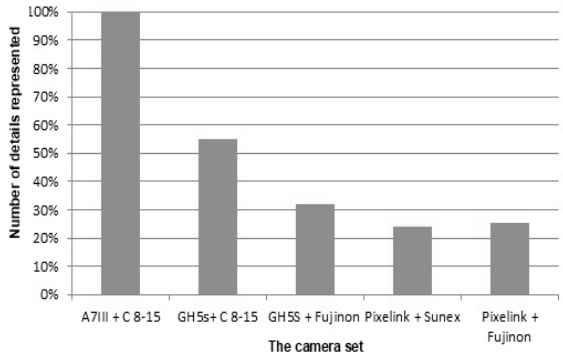


Fig. 9 *Summary of the amount of detail possible to be reproduced by a given set*

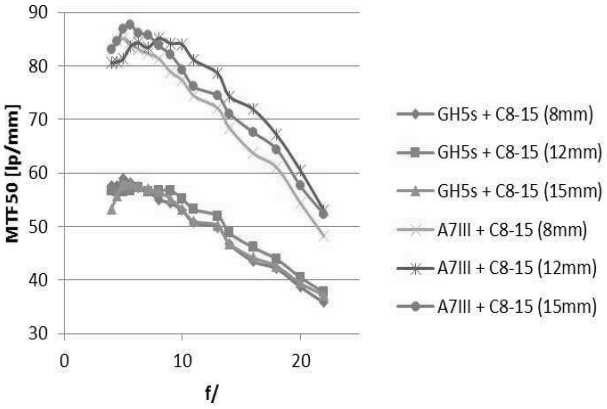


Fig. 10 *Comparison of the MTF50 value for the first and second set*

An analogous summary was made for the Pixelink + Fujinon and GH5s + Fujinon sets. The lens in both cameras was fixed without the use of additional optical elements. It enabled direct comparison of the MTF parameter in the lp/mm unit.

Both graphs (Fig. 11) are marked by a similar pattern. The average difference between the GH5S + Fujinon set and the Pixelink + Fujinon one was equal to 13% in favour of the former. It results from the use of different cameras, whose matrices have pixels of different sizes, as well as possibly differing RAW file demosaicing algorithms.

Measurement results of the lateral chromatic aberration are presented in Fig. 12. The A7III + C 8-15 set is marked by an aberration of approximately 0.05%, thus practically negligible.

Fig. 12 depicts a graph with the summary of the results for all tested sets. The Canon 8-15 and Fujinon lenses were examined twice, with different camera and adapter models. Regarding the Fujinon lens, it was predicted that the aberrations measured in different sets would have similar intensities and patterns, while with Canon 8-15, a slight increase in aberration was expected due to the use of the adapter with an additional optical element. In practice, the results obtained with the use of the GH5S camera diverge strongly from the expected values and are (about 2 to 3 times) greater than the analogous ones obtained for the remaining cameras (Fig. 13-14). The divergence is caused by the use of different RAW editing software (Adobe Camera Raw), which, despite fixed settings, probably interfered in particular image colours. The impact of the demosaicing algorithm cannot be eliminated, either. The divergence can be influenced by a different size of a single pixel of a camera matrix, yet this cannot account for such big discrepancies.

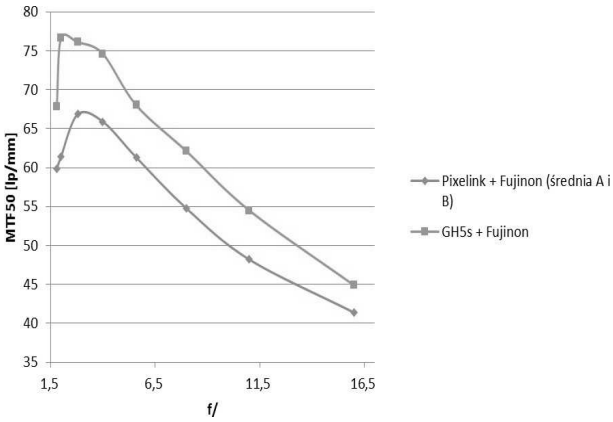


Fig. 11 *Comparison of the MTF50 value for the third and fifth set*

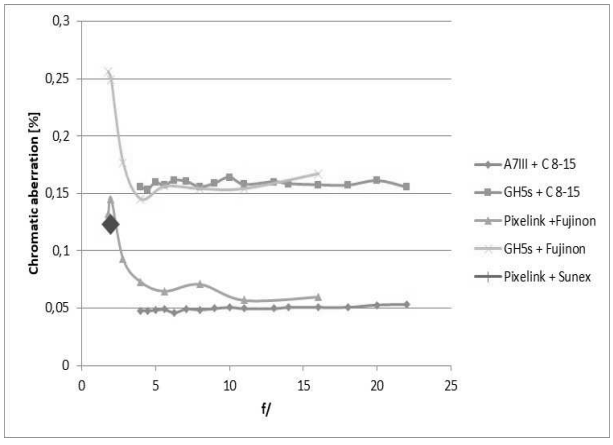


Fig. 12 *Measured chromatic aberration for camera sets*

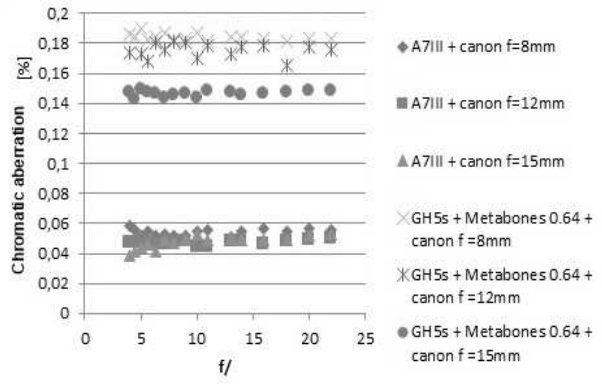


Fig. 13 *Comparison of chromatic aberration of Set 1 and Set 2*

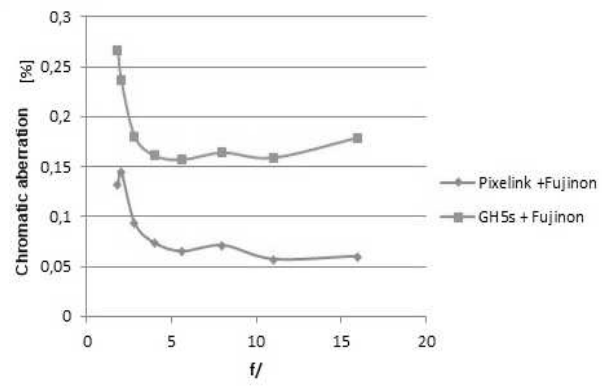
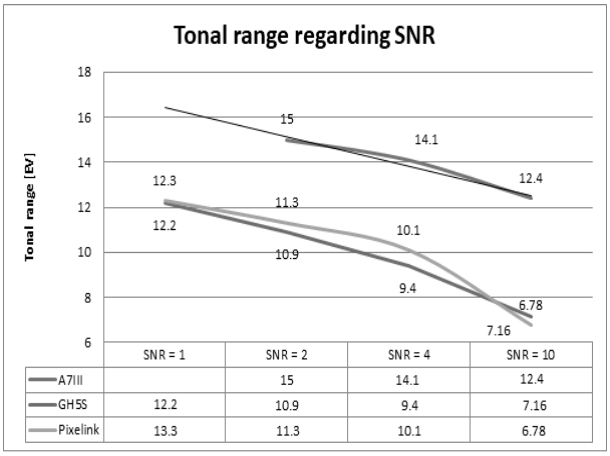


Fig. 14 *Comparison of chromatic aberration for the Fujinon lens as well as the GH5s and Pixelink cameras*

The third study area was aimed at the determination of the maximum optical resolution as well as achievable image parameters in various lighting and environmental conditions. Within this area, the investigation of the tonal range was conducted. For each camera, the tonal range was measured for four SNR values. The final result was taken as the one for SNR=1(0db) as compliant with the ISO 15739:20134 standard. The obtained results are included in Fig. 15.



	SNR = 1	SNR = 2	SNR = 4	SNR = 10
A7III		15	14.1	12.4
GH5S	12.2	10.9	9.4	7.16
Pixelink	13.3	11.3	10.1	6.78

Fig. 15 *Tonal range of cameras relative to SNR*

The greatest tonal range 15EV was obtained for the A7III camera with SNR=4. The measurement for SNR=1 failed, since below a certain exposure level on the test pattern, the value of pixel intensity for ever lower values of test fields did not fall, and the noise remained at a constant level. It is probably the effect of the limitations of the measurement method (multiple exposure of reflection pattern) for measuring wide tonal ranges. Presumably, the use of a method based on a transmission pattern would enable more accurate measurement. Using the results obtained for the remaining values, the approximation by a polynomial of degree 6 (the black line on the graph above). For SNR=1, the approximated value is close to 15EV. For the highest quality (SNR=10), A7III reaches 12.4EV, which is a very good result. For the GH5s and Pixelink cameras, the measurement for all SNR values was performed successfully. Their tonal range equalled over 12EV for SNR=1. In the highest quality (SNR=10), both cameras reached about 7EV, which is a satisfactory result. All cameras reached the tonal range of not less than 12EV. The best result was achieved by Sony A7III-15EV. The high value of this parameter predestines this camera for its use in both poor and very diverse lighting conditions. The Pixelink and Panasonic Lumix GH5S cameras are characterized by the range at the level of 12EV.

4 Example of the application in manufacturing system for the facilitation of video surveillance at work stands

Vision systems for virtual reality have wide applications not only in the entertainment industry, providing the viewer with the realistic transmission of such events as matches or concerts, but also in professional fields. There is a wide application area related to training, especially concerning work stands, where an error may bring about negative consequences for the health and life of the worker, as well as in the medical sector, where there is a restrictive limit on the number of persons admitted to the operating theatre for training purposes. VR system enables realistic transmission of an operation and facilitates various configurations of viewing angles and visualized areas depending on the needs of the people taking part in the training (specific issues depending on the specialization of the trained staff). VR vision systems are also very popular in industry. The vision system presented in the paper was designed for use in an intelligent assistance system of video surveillance and the detection of adverse events in production halls. Fig. 16 presents a schematic overview of the suggested solution.

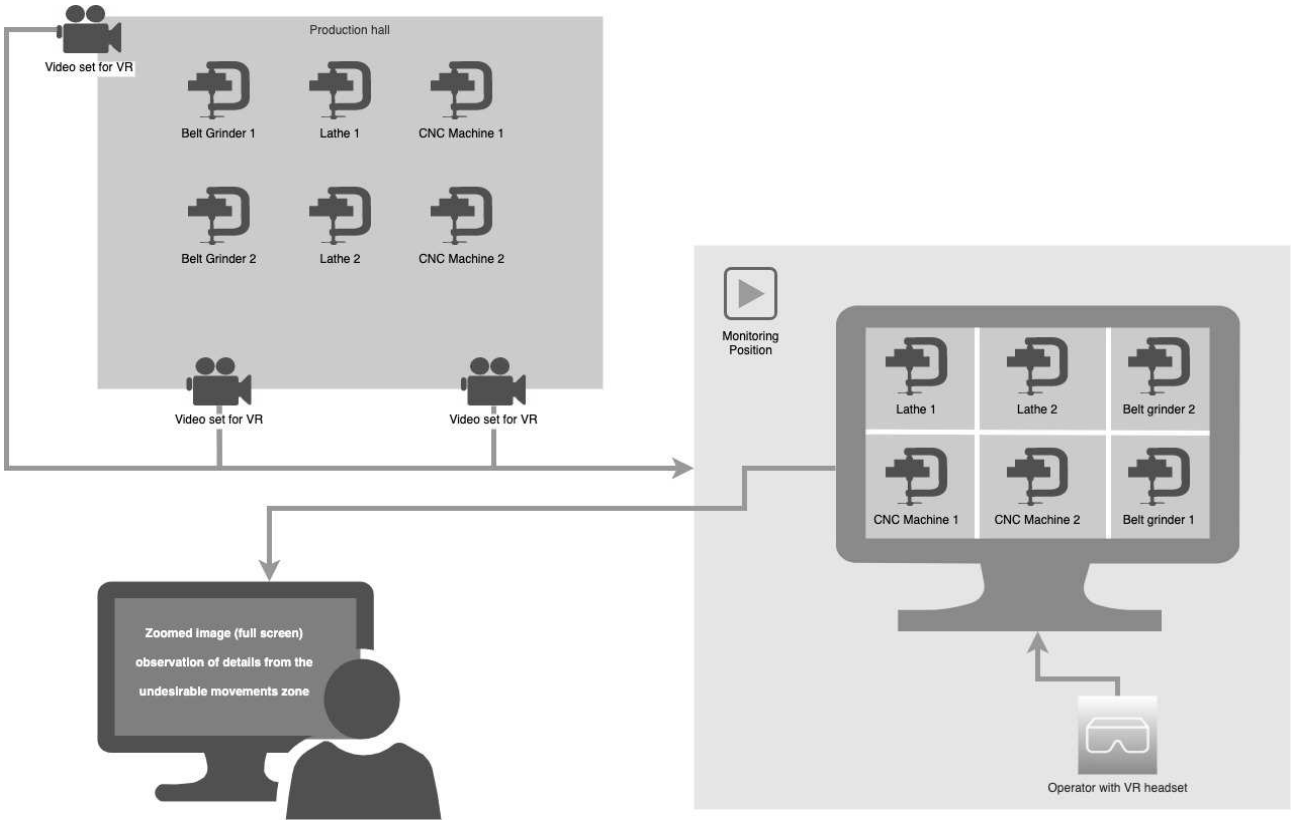


Fig. 16 *Intelligent video surveillance system in a production hall*

It is assumed that key areas such as operators' workstations or places on the production line requiring special surveillance will be filmed by means of appropriately calibrated sets for stereoscopic video. The camera set will be placed on a specially designed ring together with a device for capturing and transmitting video stream (DfCT). DfCT is an element of so-called edge computing enabling compression, correction in near real-time as well as storage and selective access to chosen video stream data. It enables an increase in the effectiveness and safety of the VR system as well as a reduction in the costs of disk space for data storage. The person responsible for video surveillance and equipped with VR glasses has a view of particular areas with the immersion effect, i.a. being present at the scene. In order to facilitate work, the system's algorithms identify undesirable movements or events and display a magnified image of the place that the alert concerns. It enables a very precise assessment of events and irregularities and their impact on the entire manufacturing process and also causes. The identification of factors that generate interference in the manufacturing process at the time of its occurrence constitutes an extremely valuable piece of information later used to redesign the process making it resistant to interference and hence more effective. It also concerns issues related to designing work stands and work optimization. The use of such a video surveillance system will be especially significant during a third shift, which is usually subject to less strict supervision procedures owing to a lower number of management staff. Lower effectiveness and a higher number of errors reported at this time also result from a lower level of human concentration. This issue could be analyzed in detail thanks to data obtained via a monitoring centre equipped with a VR system. The macro scale production system monitoring in this way is an innovative application possibility which has not been found in any literature so far. Most of the studies focus on monitoring manufacturing process in micro approach using digital twin or other algorithms specialised for the quality defects detection of particular manufactured elements wheras the security issues and safety are still to be covered especially with the growing trends of minimizing staff present in the production halls, especially during nights shifts but not exclusively.

The VR system designed, verified and tested could also serve as a training system in the fields where production process is especially dangerous for unqualified staff. In this case, the preliminary exercise in Virtual Reality would add to the safety and confidence of the operators trained later on in the real workstation. An example of this application can be a production line where the aluminium alloy is transported from the bigger furnaces to the smaller ones and then further to the production line. The temperature of the feedstock after getting out of the device reaches the temperature between 500-700

Celsius degrees centigrade.It has to be culled from the device in portions weighted 500 kg and transported to the next phase of the process realized on the three lines with 16 smaller furnaces at each of them. These devices have dispensers which administer doses of the alloy to the injection moulding machines where the subsequent steps of processing are made. For efficient realization of the process, there is a need for at least twenty well trained, experienced operators. It is extremely important for this part of the production line to strictly stick to all the instructions and safety procedures since the raw material is very hot and even small mistakes may result in deadly injury. The part of the job associated with manoeuvring the ladle to fill it up and then transport and pour the content to the smaller heating device needs special attention. The newly hired staff for this job gets training for both the theoretical and the practical skills needed. It is done with participation of well experienced operators who share information about the movements that require special attention and focus. Even though all the risks and dangers are carefully discussed and the safe work methods are presented, the management still observes inadequacies which carry a potential danger for the health of the staff. It is especially frequent with operators who work shorter than four months. Applying a new training system based on VR technology would enable more effective learning and reduce the costs of supervising the new staff. Using VR technology and stereoscopic view gives much more possibilities of confronting trainees with different situations reliably simulated in a safe way in comparison to standard video recording.

5 Summary and conclusion

The study conducted answers to the formulated problem of the view ankle and image quality. It shows that it is possible to build an effective video system for virtual reality on the basis of devices selected for research. The best test results were obtained for the first set: Sony A7 III (ILCE–7M3) + Metabones MBEF-E-BT5 adapter without optics + Canon EF 8-15mm f/4L Fisheye USM. It is predestined for being used in subsequent works on the development of a video surveillance system for adverse events/movements in production halls. The mechanical connection of the adapter had no backlash and was suitable for easy assembly and dismantling with both the camera and adapter. No impact of the use of an adapter on the image quality was found. Electrical functions, such as the settings of aperture and autofocus parameters, worked flawlessly. The adapter showed proper bidirectional communication with the camera, which was confirmed by the right reading of the fixed focal length of the lens. An additional advantage of the adapter is a camera screw jack ¼ " enabling easy attachment to a tripod in a place close to the set's centre of gravity. The greatest error relative to F-theta was equal to 6% for angles 86 and 87.5◦. Regarding the optical resolution, the obtained results are the highest for the focal length of 8 mm, and the lowest for 12 mm. The highest result was equal to 51.7 lp/mm, which translates into 2458 LW/PH (for 8 mm and f/4.5). The maximum for 8 mm is beneficial since this focal length offers a viewing angle of 180◦ used in semi-spherical images. The level of monochromatic aberration was very low, which indicates a high quality of the lens under study. In all the examined combinations, the aberration level is lower than 0.06. The dependence of aberration on focal length is visible, while the impact of the aperture is negligible. The obtained results are consistent with the subjective observations of the generated photographs – the aberration is visible only under high magnification. This brings the conclusion that the verified set of equipment (camera+adapter+lens) with addition of the special edge device equipped with the algorithms for image processing, correction and compressing provides a very reliable base for building new systems of operator's training and monitoring of the manufacturing stations and entire hall. Both concepts of implementing the multi-camera motion capture and processing system have a better potential for effective practical use due to the results shown in the study. The system evaluated with the best results provides better immersion effect and precision due to the wide angle with the smaller need for the image transposition and completion on the edges of the view area. This is especially important in a monitioring system where the VR viewer will got to the details of the undesirable movements or events taking place in the production plant. It is also important for the training system proposed since it requires precise vision of the entire working area. It is very probable that this set has numerous other applications in the VR systems developed and implemented in the manufacturing systems especially those under transformation process towards Industry 4.0 and 5.0. The latter underlines importance of safety and comfort of the human being while cooperating with higly automated and robotized manufacturing environment.

Acknowledgement

The paper is based on research from project granted by European Union and financed by European Regional Development Fund under the Smart Growth Program. The project is implemented as a part of 'Szybka Ścieżka' competition organized by The National Centre of Research and Development. POIR.01.01.01-00-1111/17.

References

[1] ESWARAN M., BAHUBALENDRUNI M. V. A. R.., (2022). Challenges and opportunities on AR/VR technologies for manufacturing systems in the context of industry 4.0: A state of the art review. In: *Journal of Manufacturing Systems*, Vol. 65 , pp. 260-278. DOI:10.1016/j.jmsy.2022.09.016.

[2] ŻYWICKI K., ZAWADZKI P., GÓRSKI F., Virtual Reality Production Training System in the Scope of Intelligent Factory. in: *International Conference on Intelligent Systems in Production Engineering and Maintenance –ISPEM*, pp. 450-458. ISSN 1854-6250.

[3] BAKAR M. R. A., RAZALI N. A. M., WOOK M., ISMAIL M. N., SEMBOK T.M.T, (2021), Exploring and Developing an Industrial Automation Acceptance Model in the Manufacturing Sector Towards Adoption of Industry4.0. In: *Manufacturing Technology Journal*, Vol. 21, No. 4, pp. 434-446. ISSN: 1213-2489.

[4] CHANG M. M. L., NEE A. Y. C., ONG S. K., (2020). Interactive AR-assisted product disassembly sequence planning. In: *International Journal of Production Research*, Vol. 58, pp. 4916–31. https://doi.org/10.1080/00207543.2020.1730462.

[5] SIEW C. Y., ONG S. K., NEE A. Y. C. (2019). A practical augmented reality-assisted maintenance system framework for adaptive user support. In: *Robotics and Computer-Integrated Manufacturing*, Vol. 59, pp. 115–29. https://doi.org/10.1016/j.rcim.2019.03.010.

[6] HALIM A.Z.A. (2018). Maintenance process in automotive industry. In: *Journal of Fundamental and Applied Science*. Vol. 10 No.3S, pp. 412-421. ISSN 1112-9867.

[7] BILLINGHURST M, HAKKARAINEN M, WOODWARD C. (2008). Augmented assembly using a mobile phone. In: *MUM'08 - : Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia*, pp. 84–7. https://doi.org/10.1145/1543137.1543153.

[8] EVANS G, MILLER J, IGLESIAS PENA M, MACALLISTER A, WINER E. (2017). Evaluating the Microsoft HoloLens through an augmented reality assembly application. In: *Proceedings Volume 10197, Degraded Environments: Sensing, Processing, and Display*; 101970V, https://doi.org/10.1117/12.2262626.

[9] ALVES J., MARQUES B., OLIVEIRA M., ARAÚJO T., DIAS P., SANTOS B.S. (2019) Comparing Spatial and Mobile Augmented Reality for Guiding Assembling Procedures with Task Validation. In: *Proce. 19th IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. https://doi.org/10.1109/ICARSC.2019.8733642.

[10] MOURTZIS D, SIATRAS V, ANGELOPOULOS J, PANOPOULOS N. (2020). An augmented reality collaborative product design cloud-based platform in the context of learning factory. In: *Procedia Manufacturing*, Vol. 45:546–51. https://doi.org/10.1016/j.promfg.2020.04.076.

[11] SCHUMANN M, FUCHS C, KOLLATSCH C, KLIMANT P. (2021). Evaluation of augmented reality supported approaches for product design and production processes. In: *Procedia CIRP*; Vol. 97, pp. 160–5. https://doi.org/10.1016/j.procir.2020.05.219.

[12] ANANTHI K, RAJAVEL R, SABARIKANNAN S, SRISARAN A, SRIDHAR C. (2021). Design and fabrication of IoT based inventory control system. In: *7th International Conference on Advanced Computing and Communication Systems (ICACCS)* pp. 1101–4.. https://doi.org/10.1109/ICACCS51430.2021.9441701.

[13] STOLTZ MH, GIANNIKAS V, MCFARLANE D, STRACHAN J, UM J, SRINIVASAN R. (2017). Augmented reality in warehouse operations: opportunities and barriers. In: *IFAC-PapersOnLine*, Vol. 50, Issue 1, pp. 12979-12984. https://doi.org/10.1016/j.ifacol.2017.08.1807.

[14] ETONAM A. K., DI GRAVIO G., KULOBA P. W., NJIRI J.G. (2019). Augmented reality (AR) application in manufacturing encompassing quality control and maintenance. In: *International Journal of Engineering and Advanced Technology (IJEAT)* (Online), Vol.9 No. 1, 197–204. ISSN: 2249-8958.

[15] URBAS U, VRABIˇC R, VUKAˇSINOVI´C N. (2019) Displaying product manufacturing information in augmented reality for inspection. In: *Procedia CIRP,* Vol.81, pp.832–7. https://doi. org/10.1016/j.procir.2019.03.208.

[16] SEGOVIA D, MENDOZA M, MENDOZA E, GONZ´ALEZ E. (2015). Augmented reality as a tool for production and quality monitoring. In: *Procedia Computer Science*, Vol. 75, pp.291–300. https://doi.org/10.1016/j.procs.2015.12.250.

[17] KOKKAS A, VOSNIAKOS GC. (2019). An augmented reality approach to factory layout design embedding operation simulation. In: *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Vol. 13, pp 1061–1071. https://doi.org/10.1007/s12008-019-00567-6.

[18] NEE AYC, ONG SK, CHRYSSOLOURIS G, MOURTZIS D. (2012). Augmented reality applications in design and manufacturing. In: *CIRP Annals* , Vol. 61, No. 2, , pp 657-79. https:// doi.org/10.1016/j.cirp.2012.05.010.

[19] OLWAL A, GUSTAFSSON J, LINDFORS C. (2008). Spatial augmented reality on industrial CNC- machines. Eng Real Virtual Real, In: *Proceedings of SPIE - The International Society for Optical Engineering* pp. 6804:680409. https://doi.org/ 10.1117/12.760960.

[20] ZHANG J, ONG S. K., NEE A.Y.C. (2010). A multi-regional computation scheme in an AR-assisted in situ CNC simulation environment. In: *Computer-Aided Design*, Vol. 42, Issue 12, pp. 1167-77. https://doi.org/10.1016/j.cad.2010.06.007.

[21] VAIDYA, S., AMBAD, P. AND BHOSLE, S. (2018), Industry 4.0 - A Glimpse, Procedia Manufacturing. *Procedia Manufacturing*, Vol. 20, pp. 233–238. https://doi.org/10.1016/j.promfg.2018.02.034

[22] KRYNKE M. Personnel Management on the Production Line Using the FlexSim Simulation Environment, In: *Manufacturing Technology Journal*, Vol. 21, 2021, No. 5, pp. 657-667. ISSN: 1213-2489.

[23] KRCMARIK D., PETRU M., MASIN I. (2019). Increasing of Precision Technology of Glass Sorting Based on Very Fast Reconfigurable Image Processing, In: *Manufacturing Technology Journal*, Vol. 19, No. 3, pp. 431-438. ISSN: 1213-2489.

[24] DOBROCKY D., JOSKA Z., STUDENY Z., POKORNY Z., SVOBODA E., (2020). Quality Evaluation of Carburized Surfaces of Steels Used in Military Technology, in: *Manufacturing Technology Journal*, Vol. 20, No. 2, pp. 152-161. ISSN: 1213-2489.

[25] SITZMANN V., SERRANO A., PAVEL A., AGRAWALA M., GUTIRREZ D., MASIA B., WETZSTEIN G. Saliency in vr: How do people explore virtual environments? [online2021-09-21].

[26] ROLDÁN J., CRESPO E., MARTÍN-BARRIO A., PEÑA-TAPIA E., BARRIENTOS A. (2019). A training system for Industry 4.0 operatos in complex assemblies based on virtual reality and process mining, In: *Robotics and Computer Integrated Manufacturing* Vol. 59 305–316. https://doi.org/10.1016/j.rcim.2019.05.004

[27] GAVISH N. , GUTIÉRREZ T., WEBEL S., RODRÍGUEZ J., PEVERI M., BOCKHOLT U., TECCHIA F. (2015) Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks, In: *Interactive Learning Environments*, Vol.23, No. 6 pp. 778–798. https://doi.org/10.1080/10494820.2013.815221

[28] GRABOWSKI A., ANKOWSKI J.J. (2015), Virtualreality-based pilot training forunder ground coal miners, *Safety Science*, Vol. 72:310–314. DOI:10.1016/j.ssci.2014.09.017.

[29] PÉREZ L., DIEZ E., USAMENTIAGA R., GARCÍA D.F. (2019). Industrial robot control and operator training using virtual reality interfaces. *Computers in Industry*, Vol. 109, No. 6, pp.114-120. DOI:10.1016/j.compind.2019.05.001

[30] WOLFARTSBERGER J., ZENISEK J., SIEVI C., (2018). Chances and limitations of a virtualreality-supported tool for decision making in industrial engineering. IFAC-PapersOnLine, Vol. 51, No. 11, pp. 637-642. DOI:10.1016/j.ifacol.2018.08.390.

[31] ANDERSON R., GALLUP D., BARRON J.T., KONTKANEN, N. SNAVELY J., HERNANDEZ C. (2016), Jump: virtual reality video. In: *Technical Papers*, Vol. 35 No. 6, pp. 1–13. ISBN: 978-1-4503-4514-9/16/12

[32] MULLER W.K., ZIEGLER R., BAUER A. (1995). Virtual reality in surgical arthroscopic training. In*: Journal of Image Guided Surgery*, Vol. 1, No. 5, pp. 288–294. doi: 10.1002/(SICI)1522-712X(1995)1:5<288::AID-IGS5>3.0.CO;2-6.

[33] MIYAMOTO K.. (1964). Fish eye lens. In: *Journal of the Optical Society of America*, Vol. 54, No. 8, pp. 1060–1061. DOI:10.1364/JOSA.54.001060.

[34] CUI Z., HENG L., YEO Y.C., GEIGER A., POLLEFEYS M., SATTLET T. (2019) Real–time dense mapping for self–driving vehicles using fisheye cameras. In: *International Conference on Robotics and Automation (ICRA) IEEE*. https://doi.org/10.48550/arXiv.1809.06132

[35] FINDEISEN M., MEINEL L., HES M., APITZSCH A., HIRTZ G. (2013). A fast approach for omnidirectional surveillance with multiple virtual perspective views. In: *Computer Science Eurocon 2013*. IEEE DOI:10.1109/EUROCON.2013.6625187

[36] LEVOY M., HANRAHAN P. (1996). Light field rendering. Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, *SIGGRAPH*, pp. 31–42https://doi.org/10.1145/237170.237199

[37] WILBURN B., JOSHI N., VAISH V., TALVALA E-V., ANTUNEZ E., BARTH A., ADAMS A., HOROWITZ M., LEVOY L. (2005). High performance imaging using large camera arrays. *ACM Transactions on Graphics*. Vol. 24, No. 3, pp. 765-776. DOI:10.1145/1073204.1073259

[38] DANSEREAU D.G., SCHUSTER G., FORD J., WETZSTEIN G. A wide-field-ofview monocentric light field camera. CVPR 2017.

[39] TANAKA K., TACHI S. (2005). Tornado: Omnistereo video imaging with rotating optics. In: *IEEE transactions on visualization and computer graphics*, Vol. 11, No. 6, pp. 614–625. DOI: 10.1109/TVCG.2005.107

[40] AGGARWAL R., VOHRA A., NAMBOODIRI A.M.(2016). Panoramic stereo videos with a single camera. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3755–3763.

[41] MATZEN K., COHEN M.F., EVANS B., KOPF J., SZELISKI R.. (2017). Lowcost 360 stereo photography and video capture. In: *ACM Transactions on Graphics (TOG),* Vol. 36 No. 4, pp.1–12. https://doi.org/10.1145/3072959.3073645.

[42] ADAM M., JUNG C., ROTH S., BRUNNETT G. (2009). Real-time stereo-image stitching using gpu-based belief propagation. VMV, pp 215–224,

[43] KONRAD R., DANSEREAU D.G., MASOOD A., WETZSTEIN G. (2017) Spinvr: towards live-streaming 3d virtual reality video. In: *ACM Transactions on Graphics*, Vol. 36, No. 6, pp.1–12, DOI:10.1145/3130800.3130836

[44] Samsung gear vr. [online 2022].

[45] KOPPAL S., ZITNICK C.L., COHEN M., BING KANG S., RESSLER B., COLBURN A. (2011). A viewer-centric editor for 3d movies. *IEEE Computer Graphics and Applications*, 31(1):20–35.

[46] GAO Z., HWANG A., ZHAI G., PELI E. (2018). Correcting geometric distortionsin stereoscopic 3d imaging. *PlOS ONE*, Vol. 13, No. 10:e0205032. .https:// doi.org/10.1371/journal.pone.0205032

[47] PELEG S., BENEZRA M., PRITCH Y. (2001). Omnistereo: Panoramic stereo imaging. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 3, pp. 279–290. DOI: 10.1109/34.910880

[48] ROLDÁN J.J., E. PEÑA-TAPIA, A. MARTÍN-BARRIO, M.A. OLIVARES-MÉNDEZ, J. DEL CERRO, BARRIENTOS A., (2017) Multi-robot interfaces and operator situational awareness: study of the impact of immersion and prediction, In: *Sensors* , Vol. 17 No. 8, pp. 1720. https://doi.org/10.3390/s17081720.

[49] PEDRAM S., P. PEREZ, S. PALMISANO, M. FARRELLY, A systematic approach to evaluate the role of virtual reality as a safety training tool in the context of the mining industry(2016).

[50] FANG H., S. ONG, A. NEE, Interactive robot trajectory planning and simulation using augmented reality, *Rob. Comput.-Integr. Manuf.* 28 (2) (2012) 227–237.

[51] CECIL J., D. POWELL, VASQUEZ D. (2007), Assembly and manipulation of micro devices – a state of the art survey, In: *Robotics and Computer-Integrated Manufacturing,* Vol. 23, No. 5, pp. 580–588. DOI:10.1016/j.rcim.2006.05.010

[52] ZHANG H.. (2017). Head-mounted display-based intuitive virtual reality training system for the mining industry, In: . *International Journal of Mining Science and Technology,* Vol. 27, No. pp. 717–722. DOI:10.1016/j.ijmst.2017.05.005.

[53] MAVRIKIOS D., KARABATSOU V., FRAGOS D., CHRYSSOLOURIS G.. (2006). A prototype virtual reality- based demonstrator for immersive and interactive simulation of welding processes, *Int. J. Comput. Integr. Manuf.* Vol. 19, No. 03, pp.294–300. ISSN 1362-3052

[54] CHRYSSOLOURIS G., D. MAVRIKIOS, D. FRAGOS, V. KARABATSOU, PISTIOLIS K. (2002), A novel virtual experimentation approach to planning and training for manufacturing processes–the virtual machine shop, In *International Journal of Computer Integrated*

Manufacturing, Vol. 15, No. 3, pp. 214–221. DOI:10.1080/09511920110034978

[55] BALCO P., BAJZÍK P., ŠKOVIEROVÁ K.. (2022). Virtual and Augmented Reality in Manufacturing Companies in Slovakia, In: *Procedia Computer Science* Vol. 201 pp. 313–320, https://doi.org/10.1016/j.procs.2022.03.042

[56] HOUSHIAR H., ELSEBERG J., BORRMANN D., NUCHTER A.. (2015). A study of projections for key point based registration of panoramic terrestrial 3d laser scan. In: *Geo-spatial Information Science*, Vo;. 18 No. 1, pp.11–31. DOI:10.1080/10095020.2015.1017913

[57] SWART D. Drawing a spherical panorama. [online].