Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 11 (2017) 1552 – 1559

Procedia MANUFACTURING

27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy

A Stereo-Panoramic Telepresence System for Construction Machines

Paolo Tripicchio[a]*, Emanuele Ruffaldi[a], , Paolo Gasparello[a], Shingo Eguchi[b], Junya Kusuno[b], Keita Kitano[b], Masaki Yamada[b], Alfredo Argiolas[c], Marta Niccolini[c], Matteo Ragaglia[c], Carlo Alberto Avizzano[a]

[a]*Scuola Superiore Sant'Anna, P.zza dei Martiri, 56100 Pisa, Italy*
[b]*R&D Unit, Yanmar Co. Ltd., Japan*
[c]*Yanmar Research Europe, Viale Galileo, 55100 Firenze, Italy*

Abstract

Working machines in construction sites or emergency scenarios can operate in situations that can be dangerous for the operator. On the contrary, remote operation has been typically hindered by limited sense of presence of the operator in the environment due to the reduced field of view of cameras. Starting from these considerations, this work introduces a novel real-time panoramic telepresence system for construction machines. This system does allow fully immersive operations in critical scenarios while keeping the operator in a safe location at safe distance from the construction operation. An omnidirectional stereo vision head mounted over the machine acquires and sends data to the operator with a streaming technique that focuses on the current direction of sight of the operator. The operator uses a head-mounted display to experience the remote site also with the possibility to view digital information overlaid to the remote scene as a type of augmented reality. The paper addresses the design and architecture of the system starting from the vision system and then proceeding to the immersive visualization.

Peer-review under responsibility of the scientific committee of the 27th International Conference on Flexible Automation and Intelligent Manufacturing

Keywords: Construction Machine; Immersive Telepresence; Stereo-Panoramic; vision head; teleoperation

* Corresponding author.
E-mail address: p.tripicchio@sssup.it

Peer-review under responsibility of the scientific committee of the 27th International Conference on Flexible Automation and Intelligent Manufacturing
doi:10.1016/j.promfg.2017.07.292

1. Introduction

Nowadays, autonomous and semi-autonomous robots are experiencing a growing success in various application domains, such as agriculture, military, search and rescue, disaster recovery, etc. According to the International Federation of Robotics (IFR), the service robot market rose by 25% between 2014 and 2015 [1].

Nevertheless, as far as the construction industry is concerned, robotic technologies have not become very popular yet. Actually, various operations that require high power and high accuracy (such as panel positioning, plumbing, material handling) are still manually performed by human workers in very ine cient and dangerous ways. At the same time, the increase of population density within urban environments is reducing the space available for construction sites. Furthermore, governments are introducing regulations aimed at both reducing the environmental impact of construction works and decreasing the amount of work related injuries. That said, it is clear that the construction industry could definitely benefit from the introduction of robotic technologies in terms of productivity, human workers safety, and environmental impact. Unfortunately, in order to successfully deploy robotic technologies in construction yards, some technical issues must be tackled and overcome. Among these issues, the most relevant is almost certainly represented by the intrinsically unstructured nature of construction sites.

Actually, construction environments are typically characterized by the combination of both indoor and outdoor spaces and by the presence of a high number of moving obstacles (i.e. human workers). Even though Building Information Modelling (BIM) is gaining popularity[2], construction yards remain highly unpredictable environments, since they lack effective and efficient systems to track the motion of workers, materials and machines within themselves. In this kind of scenario, advanced teleoperated machines represent the simpler solution that nevertheless can guarantee a major increase in terms of productivity and safety. Even though in the last decade hydraulic teleoperated machines have gained some popularity in the construction industry, these devices still lack any kind of advanced functionality, since operators normally use passive controllers (consisting of joysticks/levers and switches) to individually control each joint.

In order to improve the capabilities of this kind of machines several technologies must be implemented, such as inverse kinematics[3], closed-loop position control[4], ergonomic human-machine interfaces[5],etc. Beside these features, a fundamental functionality that state of the art teleoperated construction machines currently lack is Telepresence, i.e. the possibility to drive the machine without having to be in its proximities [6,7]. Considering that these machines are usually employed to demolish buildings [3] or to decommission nuclear power plants [8,9], the possibility to keep the human operator as far as possible from such kind of dangerous and unhealthy environments represents a significant enhancement with respect to the state of the art. Since it was originally proposed at the end of 80ies, teleoperation of backhoe and construction machines has been applied to several fields. Given the importance to assess the target distance during operations, recent works also demonstrated the need to achieve a good level of immersion through stereovision or laser-3D environment reconstruction. However, the immersive manipulation of such machinery requires the operator to access a wider field of view than those typically available in common stereographic cameras [10,11].

For all the above motivations we have investigated the design of a Mobile Robotic Telepresence(MRT)[12] system that allows the operator to immersively experience a remote working environment and control an operating working machine enhanced with a panoramic stereo camera system. This work contributes to the field by the type of camera system adopted, and by the real-time experience designed for the operator thanks to the use of Head-Mounted Display (HMD). The paper is structured as follows: first we discuss the proposed interaction design. Then we discuss the camera system followed by the visualization system comprising calibration, rendering, streaming and view augmentation. The paper is closed by conclusions and future work.

2. System and Interaction

The proposed system is sketched in Figure 1 and shown in its final form in Figure 2. Two main subsystems can be identified: a remote control subsystem wore by the human operator and a vision system directly mounted on the working machine.

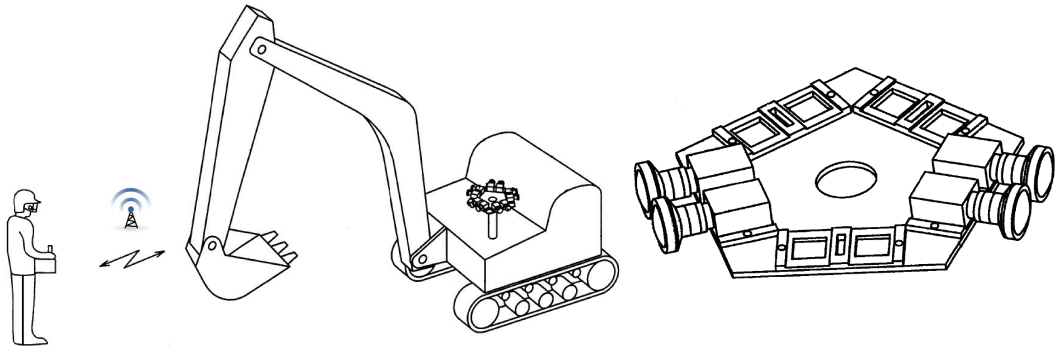

Figure 1. The human operator is shown on the left while controlling the working machine in tele-operation. The machine is depicted with the vision head mounted on top. The right part of the figure shows a detail of the vision head with two stereo pairs sketched.

These two systems are connected via wireless (WiFi 802.11ac). More in depth, the remote control system comprises a head-mounted display (HMD) and a remote control device. The display is responsible for tracking the motion of the human operator's head and for showing the scene acquired and processed by the vision system. The remote controller, instead, allows the human operator to control the motion of the working machine by means of a portable console with joysticks. On the other hand, the vision system is composed by two main elements: a multi-camera vision head and an image-processing device. The vision head comprises ten panoramic cameras arranged in pairs to be able to acquire a continuous 360□ stereo-panorama of the operating field.

This configuration of components has been chosen to allow the operator to experience the point of view of the robot while operating remotely and use its own head and body motion to verify and control the surroundings. When the operator connects to the machine or asks for a reset, the current head direction corresponds to the frontal direction of the operating machine. This reference direction is attached to the operating machine arm so that if the moving part of the machine rotates this direction rotates as well. Details about the reference frames involved in this interaction paradigm are provided in Figure 3.

Figure 2. Final system with the operator on the left and the working machine on the right. The vision head can be seen on top of the machine.

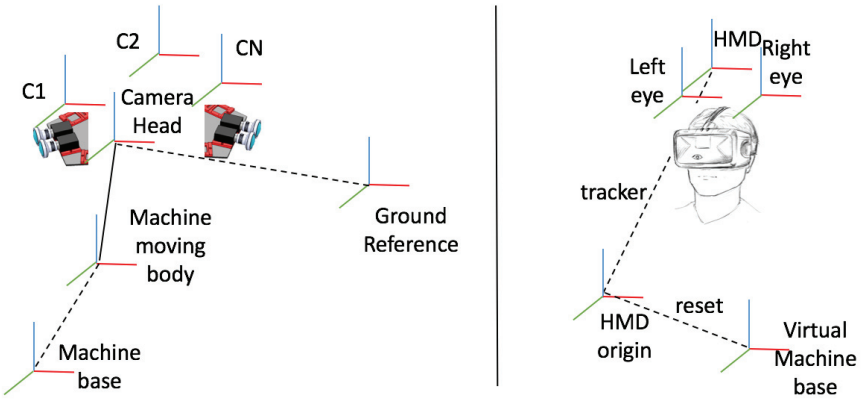

Figure 3. Reference Frames involved on the teleoperated interaction: on the left the slave site with the working machine base, attached to the swinging body over which the camera head is attached rigidly (solid line). The ground reference body used for Augmented Reality is estimated from the cameras and expressed relatively to the camera head. On the operator side on the right the operator is reported with its absolute position relatively to the HMD tracking origin. The view reset is used to match the view direction with the frontal direction of the machine.

The data exchange between the operator and the working machine allows to track the motion of the line of sight of the human operator and to accordingly adapt the image visualized on the head-mounted display in order to improve the remote control operability. Furthermore, it is worth mentioning that the vision systems guarantee a high-level of immersiveness also because the center of the lenses of the cameras are located at the "seated height" (i.e. the height of the eyes when seated) of an average human operator.

3. Camera Configuration

To provide the user with a sense of immersion and presence on the remote site, both the capture and display field of views have to be large. Here in particular we focus on the captured field of view. There are many approaches for the acquisition of a panoramic view in real-time, and they can be parametrized by the number of cameras, their layout and the field of view of each lens.

The research on omnidirectional stereo vision is much older than recent advancement in virtual reality head-mounted displays, and initial work had severe limitations in real-time performances[13], resulting in moving camera approaches. In other works, catadioptric cameras where employed for creating omnidirectional stereo vision [14].

A design requirement imposes to have a large vertical field of view (more than 135 deg), while another limits the use of camera configurations with cameras pointing around and not upward. The resulting solution is based on an circular array of fisheye cameras grouped in pairs. A fisheye lens is a system of lenses which are able to enlarge the field of view of a camera up to 190 degrees.

In our setup we employed lenses with a total 185 degrees (FE185C086HA-1) using 1" C mount. Fisheye lenses are dioptric omnidirectional cameras. Omnidirectional cameras are usually arranged by optimally combining mirrors and perspective cameras. This combination results in what is called a non-central system. In such system the optical rays are coming from the camera and get reflected by the mirror surface. These rays do not intersect into a unique point, or in a central system that exhibit the property that optical rays intersects into a unique point. A central omnidirectional camera can be built combining a pinhole camera with a mirror or using a fisheye lens. The side-effect of fisheye lenses is that image is highly distorted close to the borders and any compensation algorithm requires high resolution sensors to preserve details.

We implemented a custom design tool, coded with MATLAB (Mathworks, Natick, MA), to compare different layout configurations including different combinations of cameras and lenses. The tool takes as input the chosen cameras displacement and provides information about the field of view, the overlapping regions, and stereo capabilities for an operator observing objects at given distance.

The approach proposed in this work is based on multiple stereo-pairs that increase the sense of depth of the operator even in situation of non-immersive visualization. At the same time, the system gives the opportunity to estimate 3D depth from camera images [15]. Each pair of cameras is placed at the human average eye separation distance so that in the central part of the two cameras there are enough pixels for obtaining a stereo image after rectification, then the 5 pairs are separated by 72 degrees.

3.1. Calibration

Among the many omnidirectional camera configurations [14] we are interested in the important single viewport property for which each pixel on the sensing surface is related to a single direction of view and vice versa. This property is guaranteed by central systems in which every incoming ray passes by a single point on the mirror. An omnidirectional camera based on fisheye lenses has not the central system property but with some degree of approximation has the single viewport property, and this is sufficient for computer vision applications. The calibration process requires to find the relation between a 2D pixel point on the camera image and the corresponding 3D vector passing from the effective viewpoint. In general, the output of this process is a pair of intrinsic parameters, one for the camera and one for the mirror. To calibrate our camera system we used the OCamCalib tool [16] that uses a model that treats the imaging system as a unique compact system not caring whether there is a mirror or a fisheye lens in combination with the camera.

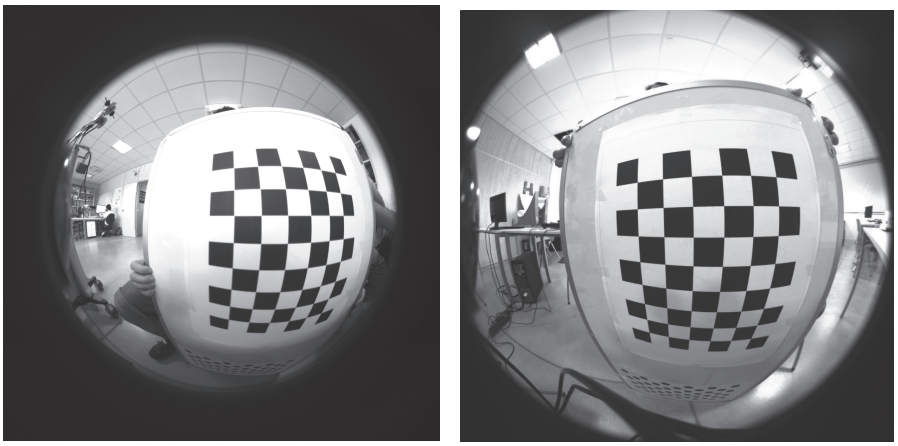

Figure 4. Example of images with chessboard used for calibrating the cameras. The left image is the full frame of the camera (2048x2048) in which camera-lens matching issues reduced the usable surface. The right image show another example with ROI (1600x1600)

The model is based on the assumption that the mirror camera system is a central system, meaning that there exist a point in the mirror where every reflected ray intersects in. This point is considered the axis origin of the camera coordinate system XYZ. The camera-calibration process uses a series of images containing a chessboard as in Figure 4 to obtain the intrinsic parameters of each Camera.

The geometric intrinsic calibration process combines a pinhole camera model with a distortion model to take into account the fisheye lenses. This allows to have both a direct and an inverse closed-form mapping between a 3D point (x, y, z) and a 2D image point (u, v). The direct mapping can be expressed in the following form:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{r\,\omega} \operatorname{atan}\left(2r \tan\left(\frac{\omega}{2}\right)\right) \begin{bmatrix} f_x & \frac{x}{y} \\ f_y & \frac{y}{z} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \tag{1}$$

where r is the radius of the normalized image coordinates point, f_x and f_y are the image focal components, c_x and c_y the image center components and ω the radial distortion.

Matching 3D points in subsequent stereo pairs it is possible to compute an extrinsic matrix between each camera pairs. If the mechanical mounting of the camera pairs is good, the extrinsic matrices should result as pure rotations along the axis passing from the center of the camera system and normal to the camera plane.

The calibration matrix is calculated as one for every camera in order to maintain a coherent reconstruction of the scene. Using the same set of cameras and lenses for each pairs, this assumption holds true. This guarantees that the image center and the points in the pair images are consistent. This property is very useful and fundamental in order to obtain a realistic stereo pair for the user eyes.

3.2. Effective field of view

The camera-lens mount introduced a 20% reduction of the effective image plate, as visible Figure 4, down from 2048 to 1600 per side. The calibration process has been repeated for each camera-lens pair using at least 12 chessboard images per camera. The effective field of view estimated from the calibration resulted in 175 degrees (from the 185 degrees of the lens).

For the purpose of visualization and also for reducing the bandwidth requirements in the acquisition system and in the network, we decided to further reduce the field of view using hardware Region of Interest (ROI) to 800x1600 pixels. After a new calibration with this frame size the effective field of view is 93 degrees horizontal and 175 degrees vertical. This means also that the horizontal field of view overlap is reduced to 21 degrees, and that horizontally we discarded regions of the images that were highly distorted.

4. Visualization System

This section presents the visualization system first discussing the calibration of the multiple-camera configuration, then the streaming system and the visualization on the operator side.

The hardware employed in the system was a Oculus HMD DK2 (94 deg of vertical field of view, 100 deg horizontal, 860 x 1080 pixels per eye). Given the vertical field of view of the HMD and the one of the camera the operator has to move upward/downward 27 degrees before encountering the upper/lower bound.

4.1. Rendering

The display of the panoramic image is based on a cylindrical projection that is placed in the virtual environment seen by the operator and located in the frame of each operator's eye (Figure 5).The virtual screen has been dimensioned so that the remote view covers the field of view of the HMD with relatively small head vertical motions (±15 deg). In particular we choose a radius of 20m, height of 100m, and is divided in 50 x 100 slices. Thanks to the above transformation each point over the cylindrical projection can be mapped to a point in the image plate, allowing a fast lookup. This mapping is performed directly by the GPU allowing to send over the network the original image from each camera pair.

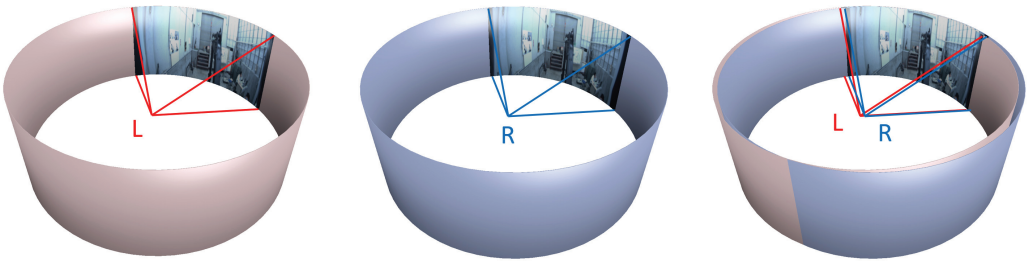

Figure 5. Concept of the Cylindrical Projection with the focus on the active Camera Pair. On the left the left-eye image projected on the cylinder is visible , in the middle the right-eye image is projected on the cylinder and at right we can see the composition of the projected images to present a correct stereo pair to the user's eyes. The separation between the cylinders is given by the user's eye separation.

The rendering of the right and left eyes is performed by associating the correct texture over the cylindrical projection wall.

4.2. Streaming

The computer that drives the system acquires the camera images at a given frequency (20Hz). At each time step only one stereo pair is selected in order to be transmitted towards the remote observer. The selection parameters and the image data are delivered on a UDP connection established in advance. The connection was implemented using the Open Source RakNet library (Oculus VR, US) due to its efficiency and NAT traversal capabilities.

The bidirectional communication involves two kinds of data. First is the stereo pair selection data. This type of data goes from the client, to the server application. The client program analyzes the head tracker data flow generated by the HMD and selects the closest stereo pair cameras based on the yaw angle. The corresponding index is sent on the RakNet connection towards the server. There, at each time step, the i^{th} stereo pair is selected and the left and right image (each 800x1600 px) are stitched together side by side in a 1600x1600 plane obtaining a stereo pair image similar to the one in Figure 6.

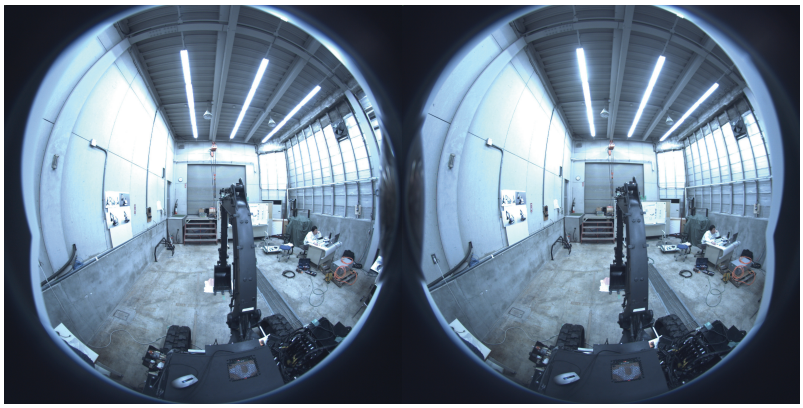

Figure 6. Example of stitched image from a stereo camera pair: this image shows the operating machine arm in front without the covers.

For the image acquisition from the USB3 cameras we opted for employing the raw Bayer format (1 BPP) format that saves bandwidth in comparison to other ones such as YUV12. This choice is at the cost of CPU for performing the conversion in a format, such as YUV12, that is suitable for video compression algorithms. We employed the fast labium library for converting RAW8 to YUV12 and then compressed the resulting image using VP8 [17]. The compressor has been configured for real-time purposes and error resiliency. It generates a constant bitrate data flow of about 5-10 Mbps depending on the configuration. Each compressed frame is delivered to the client on the same RakNet connection.

The client receives the requested stereo pair and decompress it in a YUV12 triple plane. Since the bottleneck of the client application is the bandwidth between CPU and GPU, we decided to directly upload on the GPU the YUV12 format rather than the larger RGB format (3.8MB instead of 7.6MB). A custom fragment shader has been written in order to sample the texture directly in this particular data format.

The system prototype employed a desktop computer mounted on the operating machine interfaced to the 10 USB3 cameras via special PCI eXpress hubs that allowed to access each of them at full USB3 bandwidth (625MB/s). For the chosen resolution of 800x1600 in 1 BPP at 20Hz only about 30MB/s of USB3 bandwidth are necessary considering the packet overhead.

5. Conclusions

In this paper, we have introduced and described a telepresence system for operating machines that allows the remote control of a working machine with a high level of immersivity. The wireless system is based on a pano-stereographic camera that captures the whole surrounding environment and streams the images on the operator's wearable display. The system is open at many improvements and user validation scenarios. One relevant aspect of the validation of the setup is the measure of effectiveness of the double cylinder approach provided by the fisheye stereo-pairs in comparison to a single cylinder projection with non-paired fisheye cameras. Depth cues from the environment and task specifies could be sufficient to provide depth perception even with non-stereoscopic setup [18]. We also acknowledge the importance of displaying the operator hands and remote control device. This is a good point for novice users while less problematic for expert users. The second direction of improvement is the Augmented Reality and haptic [19] feedback with specific interest in the fusion of the digital content with the 3D remote environment and the realtime robotic control of the backhoe machine.

Acknowledgements

We would like to acknowledge Yanmar R&D Europe for supporting this research. This work is covered by the International Patent WO2016079557 and related applications.

References

[1] International Federation of Robotics, Executive Summary, in: Service Robots.
[2] C.M. Eastman, C. Eastman, P. Teicholz, R. Sacks, K. Liston, BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors, John Wiley & Sons, 2011.
[3] M. Ragaglia, A. Argioas, M. Niccolini, Inverse Kinematics for Teleoperated Construction Machines: a novel user-oriented approach ,in:CIB W119 Workshop on Advanced Construction and Building Technology for Society.
[4] M. Tanzini, J.M. Jacinto-Villegas, M. Satler, C. Avizzano, M. Niccolini, An embedded architecture for robotic manipulation in the construction field, in: IEEE ETFA.
[5] M. Tanzini, J. M. Jacinto-Villegas, A. Filippeschi, M. Niccolini, M. Ragaglia, New interaction metaphors to control a hydraulic working machine's arm, in: Safety, Security, and Rescue Robotics (SSRR), 2016 IEEE International Symposium on, IEEE, pp. 297–303.
[6] L. Peppoloni, F. Brizzi, C. A. Avizzano, E. Ruffaldi, Immersive ROS-integrated framework for robot teleoperation, in: Proceedings of IEEE 3DUI, pp. 177–178. doi:10.1109/3DUI.2015.7131758.
[7] L. Peppoloni, F. Brizzi, E. Ruffaldi, C. A. Avizzano, Augmented Reality-aided Tele-presence System for Robot Manipulation in Industrial Manufacturing, in: Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology (VRST), VRST '15, ACM, New York, NY, USA, 2015, pp. 237–240. doi:10.1145/2821592.2821620.
[8] C.J. Taylor, D. Robertson, State-dependent control of a hydraulically actuated nuclear decommissioning robot. Control Engineering Practice, volume 21, Elsevier, 2013, pp. 1716–1725.
[9] R. Bogue, Robots in the nuclear industry: a review of technologies and applications. Industrial Robot: An International Journal, volume 38, Emerald Group Publishing Limited, 2011, pp. 113–118.
[10] L. Yang, N. Noguchi, Human detection for a robot tractor using omni-directional stereovision, in: Computers and Electronics in Agriculture, volume 89, Elsevier, 2012, pp. 116–125.
[11] P.-H. Yuan, K.-F. Yang, W.-H. Tsai, Real-time security monitoring around a video surveillance vehicle with a pair of two-camera omni-imaging devices, in: IEEE Transactions on Vehicular Technology, volume 60, 2011, pp. 3603–3614.
[12] A. Kristoffersson, S. Coradeschi, A. Loutfi, Advances in Human-Computer Interaction (2013)3.
[13] Z. Zhu, Omnidirectional stereovision, in: Proceedings of the Workshop on Omnidirectional Vision, Budapest, Hungary.
[14] C. Geyer, K. Daniilidis, A unifying theory for central panoramic systems and practical implications, in: European conference on computer vision, Springer, pp. 445–461.
[15] C. Ha¨ne, L. Heng, G.H. Lee, A. Sizov, M. Pollefeys, Real-time direct dense matching on fisheye images using plane-sweeping stereo, in: 3D Vision (3DV), 2014 2nd International Conference on, volume 1, IEEE, pp. 57–64.
[16] D. Scaramuzza, A. Martinelli, R. Siegwart, A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion, in: Proceedings of the Fourth IEEE International Conference on Computer Vision Systems, ICVS '06, IEEE Computer Society, Washington, DC, USA, 2006, pp. 45–. doi:10.1109/ICVS.2006.3.
[17] J. Bankoski, P. Wilkins, Y. Xu, VP8 data format and decoding guide.
[18] S. Kratz, F.R. Ferriera, Immersed remotely: Evaluating the use of Head Mounted Devices for remote collaboration in robotic telepresence, in:Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on, pp. 638–645.
[19] S.Patrinostro, M. Tanzini, M. Satler, E.Ruffaldi, A. Filippeschi, C.A. Avizzano, A Haptic-Assisted Guidance System for working machines based on virtual force fields, In Information, Communication and Automation Technologies (ICAT), 2015 XXV IEEE International Conference on. DOI: 10.1109/ICAT.2015.7340503