Augmented Reality-based Tele-robotic System Architecture for On-site Construction

Dhinesh K. Sukumar, Seongki Lee, Christos Georgoulas, and Thomas Bock

Chair of Building Realisation and Robotics, Technical University Munich, Germany
E-mail: dhinesh.sukumar@tum.de, seongki.lee@tum.de, christos.georgoulas@tum.de, thomas.bock@tum.de

ABSTRACT

The planning of construction process is crucial for successful project management and based on which all other sub-tasks and activities follow. Labour-intensive construction process which relies on field workers' own intuition and interpretation, has been too complicated to control, which can lead to mistakes, wasted time, and fatal accidents in worst but commonly happening cases. In this sense, applying innovative techniques to on-site construction process for construction planning, scheduling, operation, and monitoring has the potential to assist stakeholders in making better decisions. A head-mounted display (HMD) such as Oculus Rift can visualize construction processes information anytime whether during on-site work or beforehand by providing realistic augmented superimposed environment which is captured via the proposed 2-DOF stereo head. Moreover, by calculating the physical movement (translation, rotation) of building components, which can be accomplished by reading physical, material property data from Building Information Modeling (BIM), and combining it with computational kinematics model of on-site robotic equipment, the field workers can have the digital information for off-site component assembly in real-time manner. This research develops a new methodology for integration of tele-operation with 4D modelling in order to improve building component positioning and erection in terms of efficiency and quality by operating onsite robotic system intelligently. Based on the proposed methodology, an intelligent remotely operated, considering the control and video streaming processes, virtual tele-presence system is prototyped. The proposed system is designed for tele-operated construction robots and comprises a 2-DOF stereo head, a pair of high spatial resolution cameras, and a head mounted display with integrated head tracking mechanism. To validate the proposed methodology, controlled laboratory experiments were designed and implemented.

Keywords -
Augmented Reality; Building Information Modeling; Stereo Vision; Tele-operation; Human-Machine Interaction

1 Introduction

Safety of on-site construction is challenging while increasing productivity of the assembly process. Especially when dealing with tele-operated construction robots, the interrelated control systems should comprise a timely and robust response in order to prevent malfunctions and construction accuracy defects. Visual information plays here an important role, in case on-board mounted cameras are used in the operating space of the robot. Enhanced visual understanding of the operating space is required by the operator, especially considering robots with grippers and multi-tool end effectors. Visual information is most commonly projected into computer screens or HMDs. In case only a single camera is used to stream visual information to the HMD, the operator's perception of the optical scene is not so significant. If a pair of cameras is used, i.e. a stereo head setup, the visual perception of the optical scene is enhanced into a full 3-D visual experience [1]. The operator then can easily distinguish depth, and perform more accurate and precise pick and placement operations.

In the proposed paper, a new methodology for integration of tele-operation with 4D modelling is developed in order to improve building component positioning and erection in terms of efficiency and quality by operating onsite robotic system intelligently. Based on the proposed methodology, an intelligent remotely-operated, considering the control and video streaming processes, virtual tele-presence system is prototyped.

2 Literature Review

2.1 Robotic tele-operation

Various approaches have been proposed in the past, concerning tele-operation of robots using visual

feedback to the operators side. In [2-4], tele-robotics systems for construction robots with visual feedback are presented. Even though real-time concerns were addressed by the aforementioned approaches, the visual feedback information was projected in a single computer screen or projector panel, which doesn't allow the operator to acquire an enhanced depth perception of the operating space of the robot. Another method of robot tele-operation, which allows a human operator to remotely control a robotic manipulator is presented in [5]. It uses a contactless vision-based human-machine interface, both for the transmission of the operator motion to the robot and for the feedback of the robot back to the operator. However, the proposed thereto achieved visual feedback does not give the operator the depth visual information that is necessary for such a critical task. In [6], a highly sophisticated humanoid robot (Robonaut) is developed for space operations. It comprised a stereo head which transmits the video feedback to the operator through an HMD device. The same human-machine interface is developed also in [7] for robot control. Both implementations however, require high cost equipment and are developed based on proprietary software. In [8], web-based robot architecture is proposed, which enables the control of a robot by interacting with an advanced online graphical user interface with very promising results but the real-time constraint for control cannot be efficiently guaranteed.

2.2 Stereo Vision

Stereo vision deals with images acquired by a stereo camera setup, where the disparity between the stereo images allows depth estimation within a scene. 3-D information, hence, is retrieved which is essential in many machine vision applications such as high-speed tracking, mobile robots, object recognition and vision-guided robotics. In order to understand the structure of three-dimensional vision system it is initially necessary to describe, at least briefly, the most basic principles involved in the formation of stereoscopic image pairs. When an optical scene is observed with the human eyes, it is reflected in the retina of each eye. However, since the eyes are displaced horizontally by a specific distance, the images captured in each retina are different, as shown in Figure 1. Actually this stereoscopic pair retinas-images, contains the small displacements between the relative positions of the local parts of the image of the optical scene with respect to each image pair, depending on how close these local parts are at the point where the eyes of the observer focus. Each eye receives its own view and the two separate images are transferred to the brain for processing. When the two images arrive simultaneously, they are combined into one as shown in Figure 2.

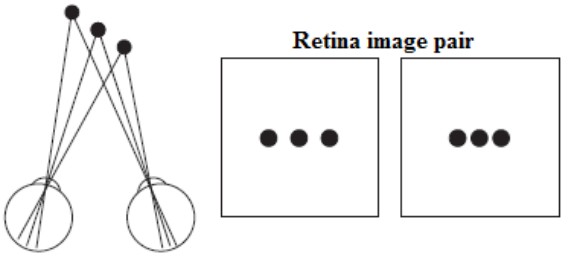

Figure 1. Due to the different perspectives, points are highlighted in slightly different positions in retinas.

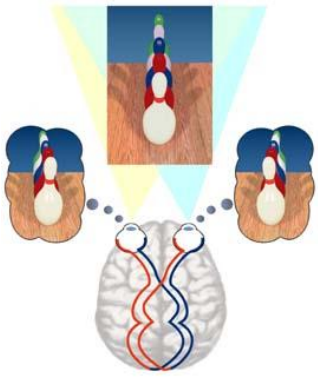

Figure 2. Fusion of individual projections for into 3-D image.

2.3 Stereo Correspondence Problem and Epipolar Geometry

Detecting conjugate pairs in stereo images is a challenging research problem known as the correspondence problem, i.e. to find for each point in the left image, the corresponding point in the right one [9]. To determine these two points from a conjugate pair, it is necessary to measure the similarity of the points. The point to be matched should be distinctly different from its surrounding pixels. Thus, in the first stage of stereo matching, suitable features should be extracted.

Two different approaches can be used to solve the stereo correspondence problem. The first approach involves the use of so-called light striping, which is a form of structured lighting. More specifically, a pattern of light strips, or an array of spotlights, project a beam of light in the field that contains the objects of the optical scene. Then process the fingerprints of the light beam to extract the information of depth. This method requires high-resolution images and a specific pattern in the light path, typically a matrix arrangement, to produce accurate results.

The second approach is the use of the epipolar lines. To understand this approach, let us assume that we have identified a specific point in the image plane of one of the stereo image pairs, and search for the corresponding point in the plane of the other image. This could be

located anywhere on the image plane, but by observing more closely at the geometry of a stereo head setup, some geometrical properties can be applied to help minimize the search field. These properties comprise the research topic of epipolar geometry [10-12]. The computational demanding task of matching can be reduced to a one dimensional search, only by accurately rectified stereo pairs in which horizontal scan lines reside on the same epipolar plane, as shown in Figure 3. By definition, the epipolar plane is defined by the point P and the two camera optical centers O_L and O_R. This plane PO_LO_R intersects the two image planes at lines EP_1 and EP_2, which are called epipolar lines. Line EP_1 is passing through two points: E_L and P_L, and line EP_2 is passing through E_R and P_R respectively. E_L and E_R are called epipolar points and are the intersection points of the baseline O_LO_R with each of the image planes. The computational significance for matching different views is that for a point in the first image, its corresponding point in the second image must lie on the epipolar line, and thus the search space for a correspondence is reduced from 2 dimensions to 1 dimension [13].

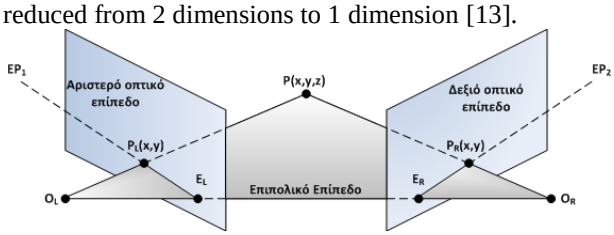

Figure 3. Geometry of epipolar plane.

2.4 Camera Calibration and Rectification

The epipolar geometry implies that the optical centres of the cameras belong to the same level, which is defined by the position of the visual scene and the optical centres of the cameras. This is necessary to have accurate results in the process of mapping. Because it is not possible to always have a stereoscopic camera setup with coplanar optical centres, a process known as camera calibration must be preceded, so we can then build on epipolar geometry.

According to the pinhole model [11], a stereo camera setup is described by a fundamental matrix, which is a representation of the intrinsic and extrinsic parameters of the camera pair. It mainly relates a point from one stereoscopic image, with a line to the other image. This line contains all the points that would correspond to this point according to the epipolar geometry. There are several ways of calculating the fundamental matrix [14-17]. In this paper, the method described in [18] was used to calculate to extract the fundamental matrix. This is a widely used technique which is based on the proposed method [17]. With the aid of a software tool, the calculation of the fundamental matrix is achieved, enabling calibration of one or more cameras. Specifically, during the process a calibration object is used which consists of a checkerboard pattern on a flat surface (Figure 4). The advantage of this method is that the calibration object may be manufactured very easily, unlike other calibration techniques which require specialized high-cost objects. With a series of images of the calibration object in different positions, (minimum two), it is possible to extract the intrinsic and extrinsic parameters of the camera, which allows the computation of the fundamental table. The fundamental matrix allows the acquired image pairs to be rectified according to the epipolar geometry, i.e. possible distortions and displacements between the two levels of the optical scenes of each of the stereo camera setup to be corrected, so that each point of an image belonging to the same epipolar line of the other.

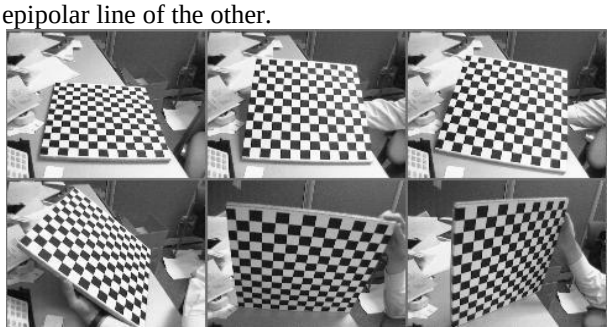

Figure 4. Images series of calibration object

In Figure 5 the image rectification is presented, following the calibration process, where aligned optical centers and epipolar lines can be observed. The red line shows the vertical displacement that exists between the optical centers of the cameras of the stereo setup. In the process of calibration and the computation of the fundamental matrix, proper (co-planar) vertical displacement is achieved, as shown by the blue line, which is now one epipolar line where the corresponding points of the left image belong to the same line in the right image.

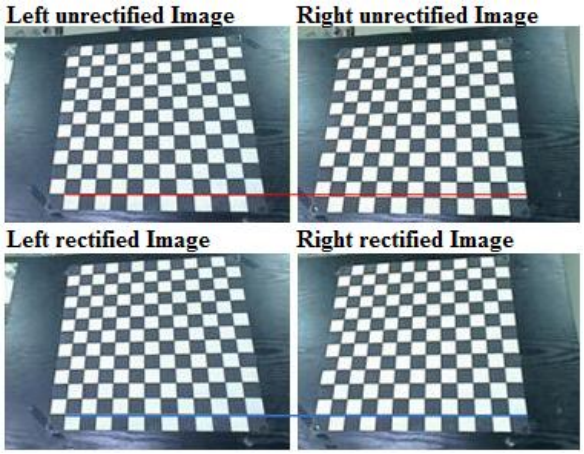

Figure 5. Stereo Image pair rectification

2.5 AR in AEC application

Many advantages of using AR system in AEC has been reported such as improvement of physical task performance and reduction of mental workload [19]. In [20] a wearable device to retrieve on-site building information is developed and validated by accomplishing shorter task completion time and higher construction correctness, which were enabled by displaying required information to on-site user. To provide accurate positioning data to the field user thorough mobile devices/AR systems, Global Positioning Systems (GPS), Wireless Local Area Network are mainly used in [21-22]. These applications are impractical for precise assembly works. Application of marker based positioning that is utilized in our research can provide cost effective and highly accurate location information even though it must be dependent on the pre installation scheme. Markers of the system are pre-attached to the scheduled construction surfaces in limited implementation size. Aligning as-is building model (3D point cloud data from the process information of stereo images) and corresponding BIM data (As-designed information) can be used to monitor the actual construction progress versus the planned schedule model.

3 Proposed System Architecture

The proposed system architecture mainly comprises the video streaming of the 2-DOF stereo camera, the control of the stereo head pan and tilt using the integrated accelerometer data from the Oculus Rift DK2 HMD, and the overlaid visual information of the Building Information Modelling regarding the construction component physical position. The implementation is performed on two dedicated computers, one on the operator side called Control Unit (CU) and the other one on the robot side called Mobile Unit (MU). The stereo head is mounted on the Motoman dual arm robot (YR-DA20-A01) on the MU side, and the HMD is connected on the CU side, where it is worn by the operator. The two host computers are currently connected together via a dedicated Wi-Fi link, to allow adequate bandwidth requirements for the video streaming process. The communication protocol between the host systems uses a higher level abstraction, built on top of UDP sockets, meeting the requirements for low latency and priority handling. In Figure 6 the overall system architecture flowchart is depicted.

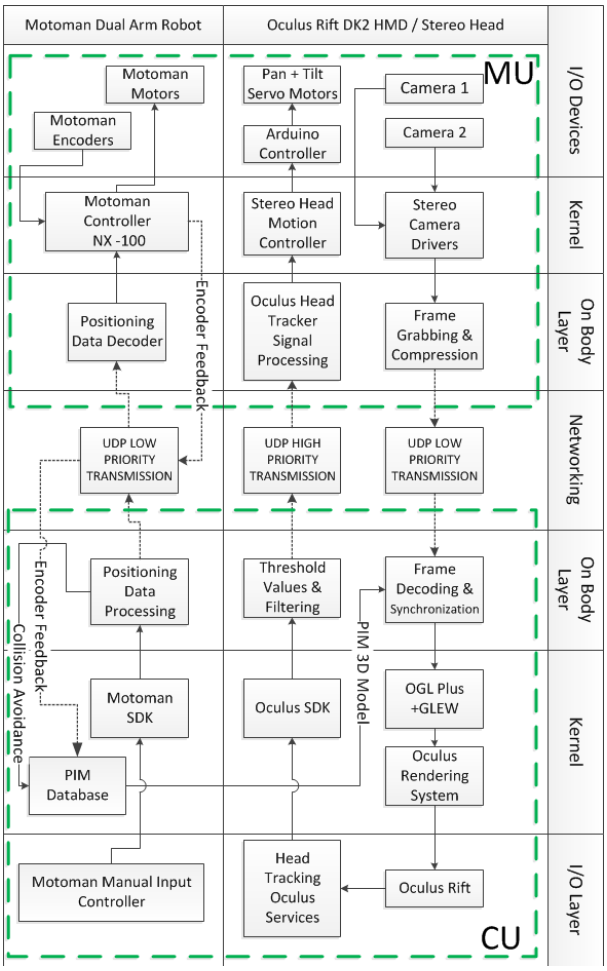

Figure 6. Proposed System Architecture Flowchart

3.1 Mobile Unit

A 2-DOF stereo head was implemented using a pair of digital servos for the pan and tilt axes. The stereo head pan/tilt rotation is driven by an Arduino micro-controller, which enabled increments of a degree, with an operating range of 0-180 degrees for each of the two axes. The stereo head was mounted on the top of the Motoman Dual Arm Robot and the relative reference positions of the stereo camera optical centres were concerned (retrieved by the stereo calibration extrinsic parameters) for the accurate overlay of the visual information to the HMD device, regarding the combined kinematics of the robot and the stereo head. The proposed designed and implemented low cost stereo head can be seen on Figure 7, and the Motoman dual arm robot with the mounted stereo head in Figure 8.

Figure 7. Implemented 2-DOF Stereo Head

Figure 8. Proposed Mobile Unit Implementation

3.2 Control Unit

The Oculus Rift DK2 HMD was used for the proposed implementation [23]. The HMD projects the left and right camera feeds from the stereo head to the user eyes, allowing to the user a 3-D perception of the optical scene. The projected camera feeds in the HMD device left and right screens can be seen in Figure 9. The HMD device embedded accelerometer sensor data are used to drive the MU side stereo head controllers. The "pitch" and "yaw" angle data from the HMD side are used to drive the "tilt" and "pan" angles respectively. With a sampling rate of up to 1 KHz the mapping between the user head rotation and the stereo head corresponding rotation allows a real-time response realization.

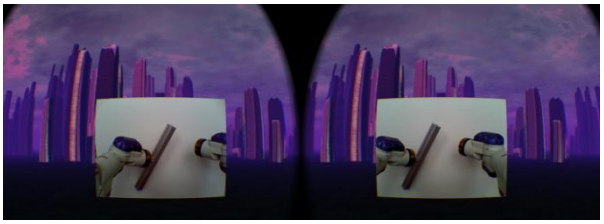

Figure 9. Oculus Rift DK2 stereo camera feeds

3.3 Building Information Model

The Building Information Model (BIM) was used for two different processes: a) to overlay into the HMD device camera feed, the actual information regarding the required position and orientation of the construction components during placing operations (in the proposed example an aluminium beam), b) to avoid possible collisions between the robot arms and the surroundings of the robot operating area.

An extension of current BIM model can describe spatio-temporal robotic assembly information, which includes direction of movement, rotation, axis, weight and time variables, among others. These attributes can for example, provide the robot in operations with minimum movement procedure to save operation energy and time.

Regarding the actual position and orientation of the construction components, a pick and place scenario of an aluminium beam which needs to be installed in a specific position on a wall was realized. The robot grabs a piece of an aluminium beam, and without a priori knowledge of the operator regarding the exact placement of the aluminium beam on the wall, the corresponding information based on the BIM data is superimposed in the HMD which is worn by the operator. Figure 10 depicts this operation, where the actual position of the aluminium beam is overlaid in red colour, and the correct position in green colour respectively. The proposed system constantly compares the BIM data, with the actual position and orientation of the robot arm end effector (this information is known from the retrieved robots' encoder data which are fed back to the model). In this way the human operator is instructed about the proper installation position of the building components, which can later manually orientate and position them accordingly based on the visual information which is projected into the HMD device.

Apart from the overlay of the proper position of the building components, a collision avoidance process is proposed. Sometimes an operator may request a specific translation or rotation of the robot arm, which might end up in a collision with the building elements within the

operating area of the robot. Once the exact position and orientation of the robot is already hardcoded into the BIM point cloud, any possible rotation motions requested by the operator are cross-checked against the BIM, and if any collisions are detected a pop up message is overlaid into the HMD device, alerting the operator accordingly.

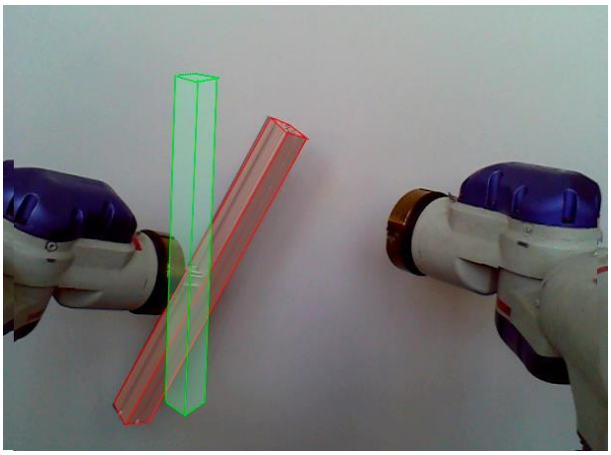

Figure 10. Projection of proper position and orientation of building component

4 Experimental Results and Discussion

4.1 FPS vs Spatial Resolution Performance

The stereo head cameras used are capable of acquiring up to 30 frames per second 1920x1080 pixels spatial resolution images. On the other side the Oculus Rift DK2 HMD device can project up to 920x1080 pixels spatial resolution images on the two on-board screens. Concerning the resource demands the high input data rate from the stereo head cameras is translated into a rate of $2(stereo) \times 1920 \times 1080(spatial\ resolution) \times 3(color) \times 8(bits) \times 30(fps) \cong 3Gbps$. Such a data transmission would undoubtedly utilize more than the available transmission bandwidth resources, considering that apart from the video stream also control commands for the stereo head and the robot need also to be stream across the communication link. Thus a compression algorithm for the video stream was used as well as a lower resolution image acquisition. The compression was implemented using the MPEG-4 codec [24]. OpenCV [25] was used alongside for the acquisition of the camera frames from the stereo head. The performed video streaming transmission rates can be seen in Table 1. For the transmission of both image frames at a spatial resolution of 920x1080 pixels (the maximum supported by the HMD device screens) the frame rate was measured at 16 fps, whereas for 460x540 pixels the proposed frame rate reached 30 fps. It was tested and noted that even a 460x540 spatial resolution setting, still results in an adequate visual feedback for the remote operator via the HMD device. An adaptive resolution selection technique was implemented to be able to swift to a higher resolution setting (decreased frame rate), when the operator requires a higher visual detail of the optical scene. To avoid eyestrain and nausea which can be caused to the operator by low frame rate and motion parallax in the optical scene, the robot as well as the stereo head rotations are de-activated.

Table 1 Video Streaming Performance

Image Spatial Resolution	Frames per Second
920x1080	16
460x540	30

4.2 BIM

This BIM framework will support the definition of relevant building product models, work procedures, and topological interaction among entities to realize the optimization of the building construction process in terms of time, money, material and safety. To this end, we need to develop automated construction surveying with BIM based 4D modelling [26] in order to improve the assembly process of building component unloading, mounting, hoisting, positioning erection, joining in terms of efficiency and quality. The building production models are represented in 4D and generated in consideration of construction engineering constraints, such as lifting capacity and speed of tower cranes, joining method and activity sequence in the long run. As a result, the 3D model of the building component is updated to reflect its actual motion in the site during assembly processes. Furthermore, by comparing the as-designed model from BIM data and the actual model of the building product from the automated surveying data, any deviations or errors between them are determined in terms of position offsets and rotation angles, which facilitate follow-up adjustment operations, which can maximize the construction quality. And BIM information management system with the functions of Information Visualization, Simulation and Optimization needs to be developed for efficient control and decision making. Moreover, by analysing and compiling building construction process patterns later on, processes configuration method will be suggested by using Web Ontology Language, Descriptive Logic (DL) and Reasoning using SWRL as suggested in [27]. On-site assembly sequences of various building construction need to be modelled case by case and will be combined into the BIM with a rule-based process configuration method.

5 Conclusions

The proposed paper describes a robust prototype construction robotics tele-operation implementation, based on the real-time stereo vision feedback and augmented building information modelling information for on-site construction applications, using low cost hardware equipment and open source software libraries. The proposed stereo head system was designed and implemented to serve as a binocular head for a remotely operated construction robot. Remote control of the stereo head via the integrated head tracker of the HMD device is achieved. The development of the real-time video streaming application of the stereo head camera pair enables enhanced depth perception experience to the operator side, allowing the precise manipulation of the remote robot.

The superimposed building information modelling information enables 4th dimension to be realized, mainly addressing two different processes: a) to overlay into the HMD device camera feed, the actual information regarding the required position and orientation of the construction components during placing operations, and b) to avoid possible collisions between the robot arms and the surroundings of the robot operating area.

References

[1] Willemsen P., Colton M., Creem-Regehr S. and Thompson M. The effects of head-mounted display mechanics on distance judgments in virtual environments, In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 35–38, 2004.

[2] Yamada H. and Muto T. Development of a hydraulic tele-operated construction robot using virtual reality - New master-slave control method and an evaluation of a visual feedback system. *International Journal of Fluid Power*, 4(2):35-42, 2003.

[3] Yamada H., Ni T. and Zhao D.X. Construction Tele-robot System with Virtual Reality. In *Proceedings of the IEEE international conference on robotics, automation and mechatronics*, pages 36-40, 2008.

[4] Tang X. and Yamada H. Haptic Interaction in Teleoperation Control System of Construction Robot Based on Virtual Reality, In *Proceedings of the 2009 IEEE International Conference on Mechatronics and Automation*, pages 9-12, Changchun, China, 2009.

[5] Kofman J., Wu X., Luu T. and Verma S. Teleoperation of a robot manipulator using a vision-based human-robot interface. *IEEE Trans. Ind. Electron.*, 52(5): 1206–1219, 2005.

[6] Bluethmann W., Ambrose R., Diftler M., Askew S., Huber E., Goza M., Rehnmark F., Lovchik C. and Magruder D. Robonaut: A robot designed to work with humans in space. *Autonomous Robots*, 14(2):179–197, 2003.

[7] Tachi S., Komoriya K., Sawada K., Nishiyama T., Itoko T., Kobayashi M. and K. Inoue, Telexistence cockpit for humanoid robot control. *Advanced Robotics*, 17(3):199–217, 2003.

[8] Marin R., Sanz P., Nebot P. and Wirz R., A multimodal interface to control a robot arm via the web: a case study on remote programming. *IEEE Trans. Ind. Electron.*, 52(6):1506–1520, 2005.

[9] Barnard S.T. and Thompson W.B. Disparity analysis of images. *IEEE Trans. Pattern Anal. Mach. Intellig.*, 2:333–340, 1980.

[10] Hartley R.I. and Zisserman A. *Multiple View Geometry in Computer Vision*, Second Edition, Cambridge University Press, 2000.

[11] Faugeras O. *Three-Dimensional Computer Vision: A geometric viewpoint*, Artificial Intelligence, MIT Press, Cambridge MA, 1993.

[12] Wexler Y., Fitzgibbon A., Zisserman A. Learning epipolar geometry from image sequences, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 209-216, Madison, Wisconsin, 2003.

[13] Georgoulas C., Sirakoulis G. And Andreadis I. *Real-Time Stereo Vision Applications*. Robot Vision, Ales Ude (Ed.), In-Tech, pages 275-291, 2010.

[14] Luong Q.T. and Faugeras O. The Fundamental Matrix: Theory, Algorithms and Stability Analysis. *International Journal of Computer Vision*, 17(1): 43-76, 1996.

[15] Hartley R.I. In defense of the 8-point algorithm, In *Proceedings of the 5th International Conference on Computer Vision*, pages 1064-1070, Boston, MA, 1995.

[16] Torr P.H.S. and Murray D.W. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3): 271-300, 1997.

[17] Zhang Z. Flexible Camera Calibration by viewing a plane from unknown orientations, In *Proceedings of the IEEE International Conference*

on Computer Vision, pages 666–673, Kerkyra, Greece, 1999.

[18] Bouguet J.Y. Camera Calibration Toolbox for Matlab, On-line: http://www.vision.caltech.edu/bouguetj/calib_doc/, 2007, Accessed: 23/01/2015.

[19] Wang X. and Dunston P.S. Compatibility issue in Augmented Reality systems for AEC: An experimental prototype study. *Automation in Construction*, 15(3):314-326, 2006.

[20] Yeh K.C, Kang M.H. and Kang S.C. On-site building information retrieval by using projection-based Augmented Reality, *Journal of Computing in Civil Engineering*, 26(3):342-355, 2012.

[21] Irizarry J., Gheisari M., Williams G., and Walker B.N. InfoSPOT: A mobile Augmented Reality method for accessing building information through a situation awareness approach. *Automation in Construction*, 13:11–23, 2013.

[22] Shin D.H. and Dunston P.S. Identification of application areas for augmented reality in industrial construction based on technology suitability. *Automation in Construction*, 17(7): 882-894, 2008.

[23] Desai1 P.R, Desai P.N, Ajmera K.D. and Mehta K. A Review Paper on Oculus Rift-A Virtual Reality Headset, *International Journal of Engineering Trends and Technology (IJETT)*, 13(4):175-179, 2014.

[24] Richardson I.E. *H.264 and MPEG-4 Video Compression*. New York, Wiley, 2003.

[25] Bradski G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[26] Hartmann T., Gao J. and Fischer M. Areas of application for 3D and 4D models on construction projects, Journal of Construction Engineering and Management 134(10):776-785, 2008.

[27] Benevolenskiy A., Roos K., Katranuschkov P., and Scherer R.J. Construction processes configuration using process patterns. Advanced Engineering Informatics 26(4):727-36, 2012.