INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

Immersive Robot Teleoperation Using an Hybrid Virtual and Real Stereo Camera Attitude Control

João José Gomes Reis

Thesis submitted to the Department of Electrical and Computer Engineering in partial fulfilment of the requirements for the degree of

Master of Science in Electrical and Computer Engineering

Panel

President: Doutor Carlos Filipe Gomes Bispo
Advisor: Doutor Rodrigo Martins de Matos Ventura
Examiner: Doutor Alexandre José Malheiro Bernardino

October 2012

Abstract

This work addresses the problem of providing effective situation awareness to the operator of a remote vehicle. Usually teleoperated systems are performed through a Graphical User Interface (GUI) that provides robot sensor's data to the operator, overloading him with information. In this work is proposed a different setup. It comprises a Head Mounted Display (HMD) worn by the operator. The HMD is equipped with a head tracker, providing the attitude of the operator head. The robot is equiped with a stereo camera mounted on a Pan and Tilt Unit. The images captured by the stereo camera are streamed to the HMD, providing 3D vision to the operator. The main goal of this setup is to stream to the HMD images with the same attitude than the one of the operator head. This is accomplished by an hybrid approach by combining the PTU orientation control through the servos with a virtual pan, tilt and roll rectification method. This rectification introduces roll rotation of the images, not possible since the PTU does not possess that degree of freedom, and overcomes the PTU servos limitation in term of responsiveness and accuracy. This arrangement allows a more intuitive camera orientation control for the operator by decoupling the camera orientation from the robot orientation and a more immersive experience through the HMD. A user study was performed to compare this system with the traditional GUI teleoperation method and with a system where the stereo camera orientation was provided directly with the HMD attitude sensor.

Keywords: Virtual pan, tilt and roll, Imersive teleoperation, 3D Vision, Hybrid pan & tilt unit control, Image projection through a new orientation, decouple of stereo camera kinematics

Resumo

Esta tese aborda o problema de aumentar o *situational awareness* do operador de um veículo remoto. Normalmente a teleoperação de um robô é realizada através de uma *Graphical User Interface* (GUI), um ecrã, onde o operador encontra toda a informação capturada pelos sensores do robô, sobrecarregando-o com informação. Nesta tese é proposto um sistema diferente para realizar a teleoperação. Este sistema é composto por um *Head Mounted Display* (HMD) usado pelo operador. O HMD tem um sensor de atitude de modo a fornecer a atitude da cabeça do operador. O robô está equipado com uma câmara stereo montada numa unidade *pan and tilt* (PTU). As imagens capturadas pela câmara são enviadas para o HMD possibilitando a visão em 3D. O principal objectivo deste sistema é enviar imagens para o HMD com atitude igual à da cabeça do operador. Isto é realizado através de uma abordagem que combina o controlo dos motores do sistema PTU com um método de rectificação virtual de *pan, tilt and roll.* Este método introduz o ângulo de *roll* que o sistema PTU não consegue realizar, porque não tem esse grau de liberdade, e ultrapassa também as limitações em termos de resposta e precisão dos motores. Este sistema providencia também o desacopulamento do controlo da orientação da câmara em relação á orientação do robô e uma experiência mais imersiva de teleoperação através do HMD. Foi realizado um caso de estudo para comparar o sistema desenvolvido com o método de teleoperação através de uma GUI e com um método em que o controlo da orientação da câmara stereo era feito directamente a partir da orientação fornecida pelo sensor de atitude do HMD.

Palavras-chave: Pan, tilt and roll virtual, Teleoperação imersiva, Visão 3D, Controlador híbrido para unidade pan & tilt, projecção de imagem segundo uma nova orientação, desacopulamento do controlo da camera stereo

List of Figures

List of Tables

Contents

Chapter 1

Introduction

This chapter is an introduction to the work developed for this thesis. In section(1.1), is presented the problem which was the starting point for this thesis. Section (1.2) follows up with a literature review and in section (1.3), are stated the main contributions of this thesis. The last section, (1.4), presents the thesis outline.

1.1 Problem Statement

The utilization of Search And Rescue robots (SAR) in danger scenarios is becoming a more and more common pratice with the advance that this area have faced trough the last decade. These robots are developed to enter inaccessible and dangerous spaces or scenarios, like collapsed buildings or radioactive areas. Their missions can go from search for victims in dangerous areas, to area reconnaissance providing crucial information of the scenario status. Usually, SAR robots are teleoperated by a skilled operator. The most common setup to provided to the operator to operate the robot is comprised by a computer and a joystick or game pad. In the computer screen, the operator sees real-time 2D images provided by the robot's cameras among all the others sensors information (atmosphere composition, infrared readings, radioactive levels, . . .). Through the joystick or game pad the operator is able to control the robot motion. The operator task is quite difficult: besides paying attention to all the sensor data, he must fully understand the environment around the robot and the consequences of his decisions concerning the robot motion.

In order to enhance the operator perception of the environment it's proposed a 3D vision system. This system is comprised by an Head Mounted Display that provides 3D vision to the operator. The 3D images are captured by a stereo camera mounted on a pan and tilt system. The HMD contains an head tracker, providing the operator head attitude. With this setup the pan and tilt system will follow the head attitude. This setup have two main advantages: the 3D vision that grants a better perception of the environment and a more intuitive control of the camera orientation. Studies have shown that this setup enhance the operator situational awareness, granting a more safe control of the robot [1]. The main problem of this setup is the pan and tilt motor responsiveness. This response tends to be slow for large changes in the camera orientation, which results in a lagged experience for the operator. Also, generic pan and tilt systems have low resolution step motors, that may lead the operator to experience some tremble in the image, for small pan and tilt movements.

Figure 1.1: Vuzix Warp 920 VR Head Mounted Display, in detail the attitude sensor.

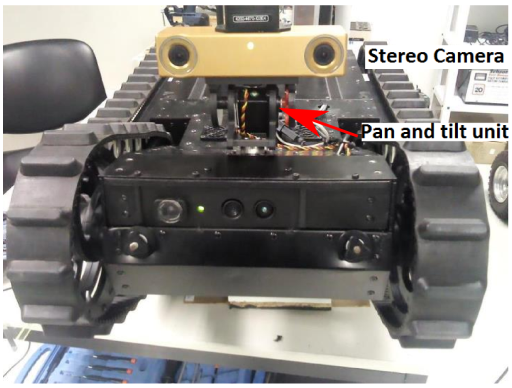

Figure 1.2: RAPOSA-NG with the stereo camera mounted on the Pan and Tilt Unit.

To overcome these limitations this thesis introduces a virtual pan, tilt and roll method to be applied in the stereo images. This method consists in projecting the images captured by the stereo camera through a new orientation, that corresponds to the difference between the operator head attitude and the camera orientation. Projecting the images through this orientation provides the operator with the right perspective instead of the camera perspective that, due to the physical limitations of the pan and tilt system, may not be correct. To take full advantage of the virtual pan, tilt and roll method it was also designed and implemented a controller that combines the method advantages with the pan and tilt system orientation.

1.2 Related Work

1.2.1 Robots Teleoperation

Robots teleoperation have been widely researched. In Henrique Martins master thesis [1] he implemented a 3D teleoperation system on RAPOSA with an HMD. The resuts achieved showed that 3D vision enhances the operator perception of the environment around the robot. In his work the cameras that provided the visual information to the HMD where in the front of RAPOSA meaning that the operator needed to rotate the robot to see the environment. He concluded that with the 3D vision through the HMD the operator was able to better perceive the obstacles and distance between the robot to obstacles on the robot path. The HMD influence have been studied showing a increase of the operator performance [2]. The U.S. military forces have been requesting studies in the utilization of HMD and 3D vision through flat screens to enhance robots teleoperation [3, 4].

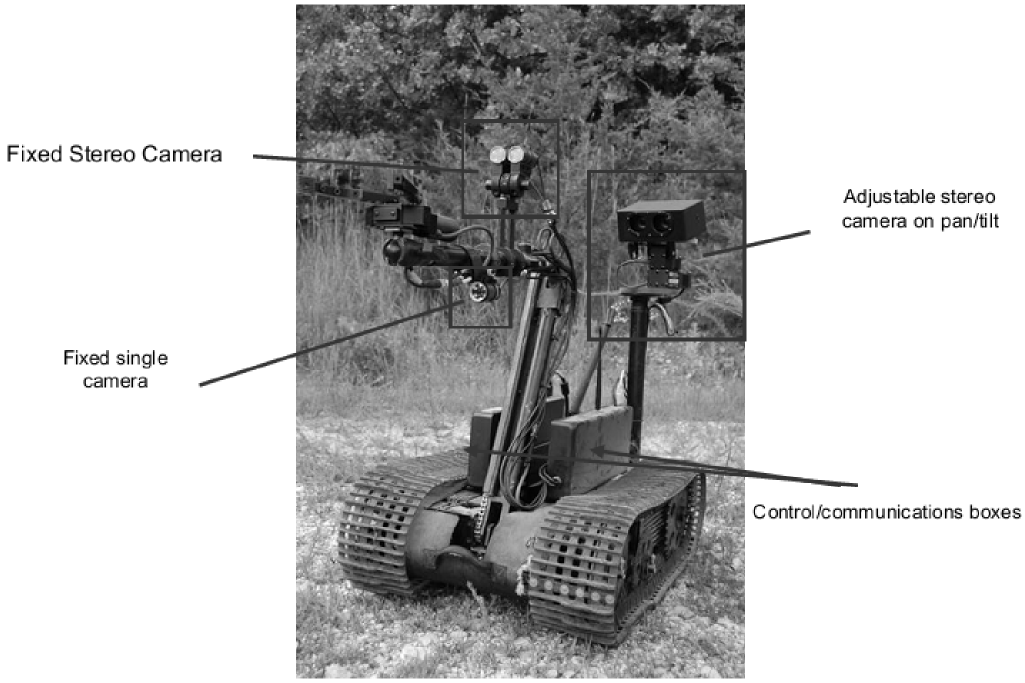

Figure 1.3: U.S. Army TALON robot with additional hardware.

1.2.2 Virtual pan and tilt

Virtual pan tilt and zoom camera simulation is a well studied subject[5, 6]. This is usually done to provide simultaneous operators to explore a wide area captured by a high resolution camera or to help in surveillance systems. The virtual pan tilt and zoom camera simulation uses a Region Of Interest smaller than the image captured a high resolution camera. This ROI is displayed to the user enabling him to perceive an area of the scenario with more detail. Providing the operator with the ability to control the position of the ROI in the image (with a gamepad, joystick or keyboard) and the operator perceives pan and tilt camera movement across the scenario. If the user can control the size of the ROI he perceives a zoom in/out effect. Although this approach provides good results for wide capture areas from a distant place, it fails in providing the right perspective if the scenario is not far enough. An approach to solve the problem is to keep a panoramic view (mosaic) of the world to display to the operator [7]. To achieve this the camera is constantly moving to update all the areas of the mosaic. The main problem with this approach is to define what are, at each moment, the more important areas of the mosaic to update that can be divided in two criteria: how long an area of the mosaic is not updated and is there new information to update an area (For example if the robot turns to a side there will be new information that the operator will want to see).

In 1990 Apple inc. developed Quicktime VR [8], a file format that allowed the creation of panoramas and the exploration of objects and scenes through images captured multiple times from different angles. This new technique provided a good view of the scenario but could only be accomplished off-line.

1.3 Contributions

This thesis presents a Virtual pan tilt and roll method developed to obtain a more immersive teleoperation experience to a robot operator in order to increase his perception of the environment around the robot. This result in a more safe control of the robot that may lead to a more successful result in the task the robot operator have to perform. A study to compare this method with the GUI traditional control method was performed with good results.

1.4 Thesis Outline

This thesis goes as follows: Chapter (2) follows up with explaining the system design and implementation, covering the mathematical background of the virtual pan, tilt and roll method and the thought process behind the controller. Chapter(3) describe the experience designed to test the implemented system and the results. Chapter(4) presents the conclusions of this work and a propose developments and suggestions for future work.

Chapter 2

System Architecture

This chapter describes the architecture of the controller designed to overcome the problem stated in section (1.1). The first section, (2.1), provides a global overview of the system.The next two sections go in with more detail: section (2.2) goes through the mathematical background and the method used to accomplish virtual pan, tilt and roll, while section (2.3) explains the process behind physical pan and tilt unit controller design.

2.1 Global Overview

The system setup is comprised by an Head Mounted Display (HMD), a Pan and Tilt Unit (PTU), and a stereo camera. The HMD have an attitude sensor and provides the user head attitude. The stereo camera is mounted on the PTU enabling to control its orientation. This architecture is presented in figure 2.1 .

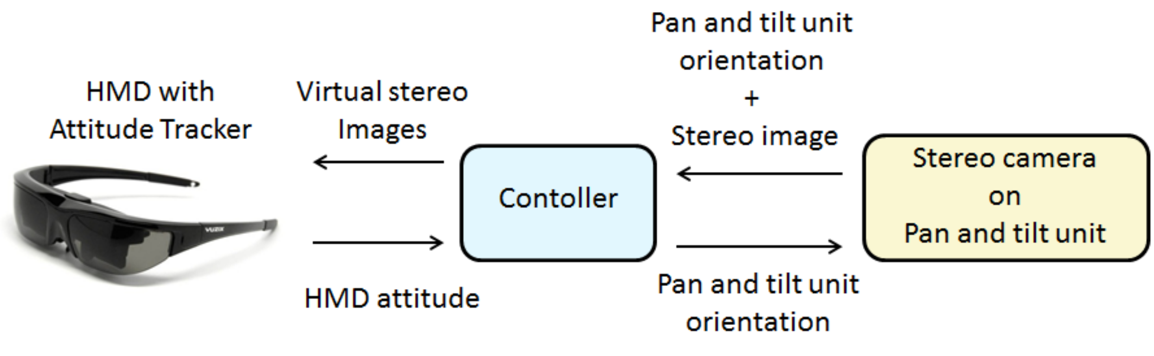

Figure 2.1: System Architecture with the Controller block

As one can see in figure 2.2, the controller block can be divided in tow smaller blocks to ease the explanation process: the (1) VPTR block where the virtual pan, tilt and roll method is applied to the stereo images from the stereo camera and the (2) pan and tilt unit controller which, as the name suggests, provides the orientation to the pan and tilt unit. Both blocks will be explained in detail over the next two sections.

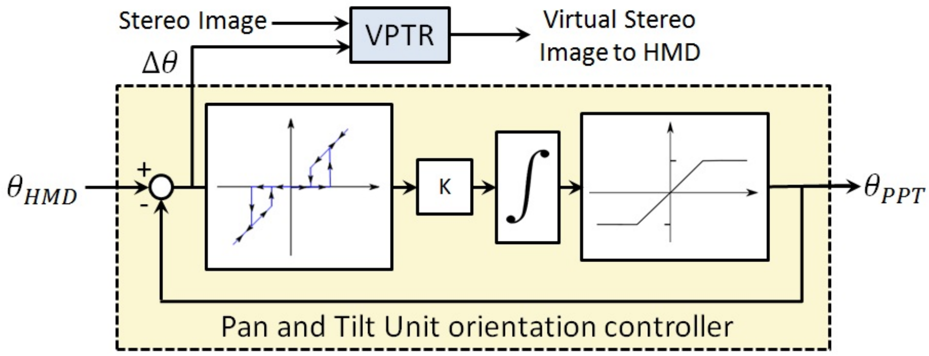

Figure 2.2: Block diagram of the controller.

Let Θ_{HMD} be the HMD attitude and Θ_{PTU} the PTU orientation in Roll, Pitch, Yaw representation (rpy) . One can define the error $\Delta\Theta$ as

$$\begin{aligned} \Theta_{PTU} &= (0, p_{PTU}, y_{PTU})^{\intercal} \\ \Theta_{HMD} &= (r_{HMD}, p_{HMD}, y_{HMD})^{\intercal} \\ \Delta\Theta &= \Theta_{HMD} - \Theta_{PTU} \end{aligned} \tag{2.1}$$

where r, p and y are the respective roll, pitch and yaw angles for each device. Since the PTU only have two degrees of freedom, pan and tilt, r_{PTU} angle is always zero.

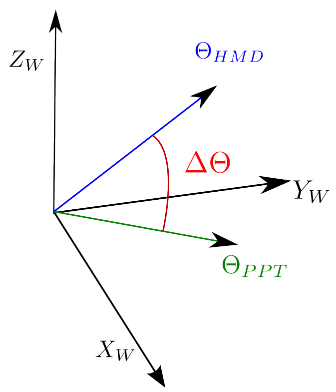

Figure 2.3: Representation of Θ_{HMD} and Θ_{PTU} in the world frame apart a translation and the difference $\Delta\Theta$

2.2 Virtual Pan, Tilt and Roll module

In this section comes the mathematical background for the virtual pan, tilt and roll method. It starts by introducing the pin-hole camera model in order to achieve back projection from the image. From there the relation between the virtual and the real camera pixels position is achieved.

2.2.1 Pinhole Camera Model

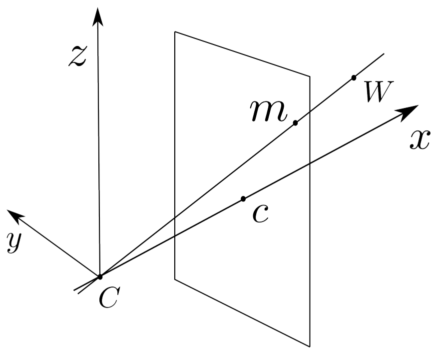

Figure 2.4: Pin Hole camera model

Using the pinhole camera model (figure 2.4), a point in the world $w = (x, y, z)^\intercal$ is projected in an image $m = (u, v)^\intercal$ in homogeneous coordinates, $\tilde{m}$ by

$$\lambda \tilde{m} = \tilde{P} \begin{bmatrix} w \\ 1 \end{bmatrix} = \tilde{P}\tilde{w} \tag{2.2}$$

The matrix $\tilde{P}_{3\times4}$ is the perspective projection matrix and is given by

$$\tilde{P} = A \begin{bmatrix} R \mid t \end{bmatrix} \tag{2.3}$$

Where $R_{3\times3}$ and $t_{3\times1}$ are the extrinsic camera parameters, the relation between the camera frame and the world frame. The intrinsic camera parameters matrix, $A_{3\times3}$ is given by

$$A = \begin{bmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.4}$$

where

α_u	Focal length in terms of pixels through x axis,
α_v	Focal length in terms of pixels throught y axis,
γ	Skew coefficient between x and y axis ,
(u_0, v_0)	Coordinates for the center of the image.

The projection matrix $\tilde{P}$ can be written as:

$$\tilde{P} = \begin{bmatrix} {q_1}^\intercal \mid q_{14} \\ {q_2}^\intercal \mid q_{24} \\ {q_3}^\intercal \mid q_{34} \end{bmatrix} = \begin{bmatrix} Q \mid q \end{bmatrix} \tag{2.5}$$

The null space of $\tilde{P}$ can be spanned from c, the camera focal point. Through 2.2 and 2.5

$$0 = \begin{bmatrix} Q \mid q \end{bmatrix} \begin{bmatrix} c \\ 1 \end{bmatrix} \tag{2.6}$$

q is given by,

$$q = -Qc \tag{2.7}$$

Therefore $\tilde{P}$ can be written has

$$\tilde{P} = \begin{bmatrix} Q \mid -Qc \end{bmatrix} \tag{2.8}$$

With this new perspective projection matrix $\tilde{m}$ is given by

$$\lambda \tilde{m} = \begin{bmatrix} Q \mid -Qc \end{bmatrix} \begin{bmatrix} w \\ 1 \end{bmatrix} \tag{2.9}$$

$$\lambda \tilde{m} = Qw - Qc \tag{2.10}$$

From this result the optical ray that connects c to any point in the world can be parameterized, apart a constant λ by

$$w = c + \lambda Q^{-1} \tilde{m} \tag{2.11}$$

There are infinite points existing in this line and they are all projected in m. The parameter λ is a positive scale factor and defines position of the 3D point in the line.

2.2.2 Virtual Pan, Tilt and Roll

The virtual pan, tilt and roll method objective is to simulate a camera rotation over the optical center (figure 2.5).

The real camera, with index rc, have a perspective projection matrix P_{rc} and the virtual camera, with index vc, have a perspective projection matrix P_{vc}. In the previous section, equation 2.11

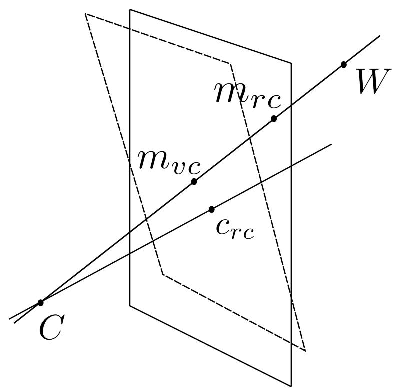

Figure 2.5: Representation of the real camera (rectangle in full line) and the virtual camera (rectangle in dashed line). The virtual camera is a rotation of the real camera over C, the optical center. The indexes rc and vc stand for real camera and virtual camera respectively.

matches a point m in the image, to a point w in the world. If both cameras are capturing the same point w, then for each camera the 3D position of w is given by

$$\begin{cases} w = c_{rc} + \lambda_{rc}\, Q_{rc}^{-1}\, \tilde{m}_{rc} \\ w = c_{vc} + \lambda_{vc}\, Q_{vc}^{-1}\, \tilde{m}_{vc} \end{cases} \tag{2.12}$$

Since the relation between both cameras is a rotation over the optical center, $c_{rc} = c_{vc}$. With w being the same for both cameras then, by matching both equations, one can write

$$\tilde{m}_{vc} = \lambda\, Q_{vc}\, Q_{rc}^{-1}\, \tilde{m}_{rc}, \quad \lambda = \lambda_{vc}/\lambda_{rc} \tag{2.13}$$

This equation matches a point projected in the real camera image to a point in the virtual camera image. From 2.3 and 2.5 one can see that for both virtual and real cameras that

$$\begin{cases} Q_{rc} = A_{rc}\, R_{rc} \\ Q_{vc} = A_{vc}\, R_{vc} \end{cases} \tag{2.14}$$

Since the objective is to simulate a rotation with the same camera, the intrinsic parameters for both cameras are the same, $A = A_{rc} = A_{vc}$. If the world frame and the real camera frame are the same then R_{rc} is an identity matrix and therefore (2.13) becomes

$$\tilde{m}_{vc} = \lambda\, A\, R_{vc}\, A^{-1}\, \tilde{m}_{rc} \tag{2.15}$$

where R_{vc} represents the rotation between the real camera and the virtual camera. Given a rotation matrix between virtual and real cameras, equation 2.15 relates the real image pixel coordinates with the virtual image pixel coordinates. As this method projects the image through a new orientation provided by the rotation matrix, the image perspective changes, and the user perceives like he is moving the real camera orientation. This method is applied in the VPTR

(a) Reference Image.

(b) Image taken with 22 degrees pitch angle.

Figure 2.6: Images taken to demonstrate the virtual pan, tilt and roll method (Note: the color of the images is different due to the camera automatic color rectification)

block to the images provided by the stereo camera. The rotation matrix is calculated from the error $\Delta\Theta$, providing to the operator the correct perspective concerning his head attitude. To test this method, two camera images were taken with different angles (figure 2.6).

The virtual pan and tilt method was applied to the reference image with a $\Delta\Theta$ of 22 degrees for the pitch orientation. The detailed result can be seen in figure 2.7.

Figure 2.7: Virtual pan, tilt and roll comparation: In the left is the detail of the image taken with a 22^o camera orientation. On the right is the result of applying the virtual pan, tilt and roll method to the reference image.

As one can see the virtual pan, tilt and roll method replicates the angles from the image taken with 22^o orientation. The main problem with this method is the image resolution. In figure 2.8 one can see the source pixels and the destination when applying the method.

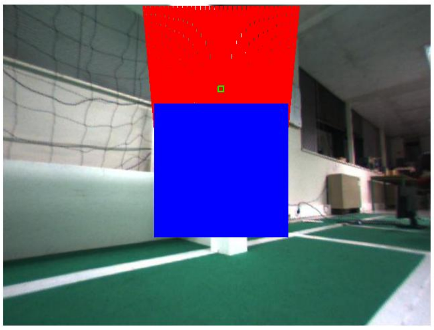

Figure 2.8: Origin of the pixels used in the method. In red are the original pixels position and in blue their destination.

2.3 Physical pan and tilt system controller

In the previous section it was introduced the VPTR block that implements the virtual pan, tilt and roll method. In this section the pan and tilt unit controller is presented. The goal is to blend both virtual pan, tilt and roll with the pan and tilt unit movement to provide the right perspective to the operator and surpassing the problems stated in section (2.1).

2.3.1 Controller design

The PTU controller block diagram is presented in figure 2.9. It comprises a cascade of blocks: a hysteresis block, a proportional block K, a saturation block and an integrator.

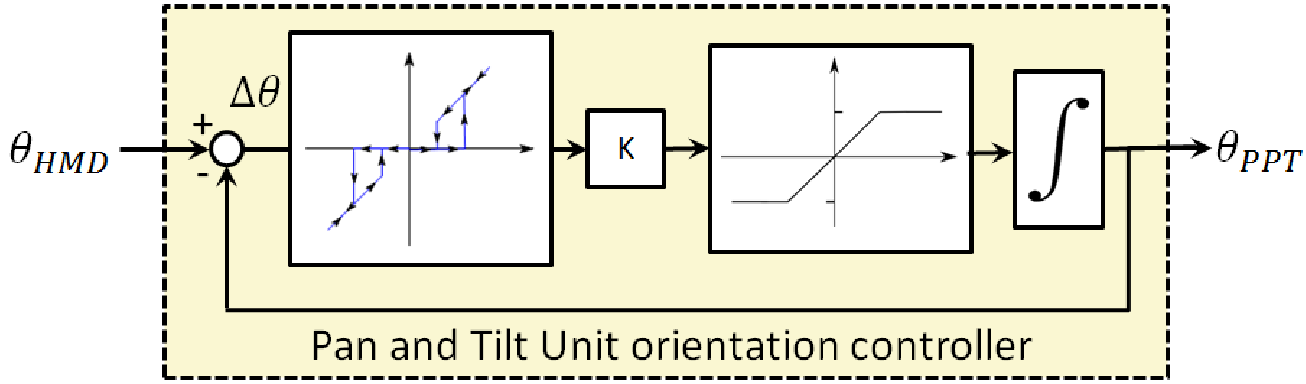

Figure 2.9: Physical pan and tilt controller

Since the pan and tilt unit only performs movement in pan and tilt angles, the error $\Delta\Theta$ will be defined only for pitch and yaw angles

$$\Delta\Theta_{py} = (\Delta\theta_p, \Delta\theta_y) \tag{2.16}$$

The first block is the hysteresis block. This block applies a non-linear transformation in the error $\Delta\Theta_{py}$ in order to provide two work modes for the pan and tilt angles: the (1) active mode and the (2) idle modes (figure 2.10).

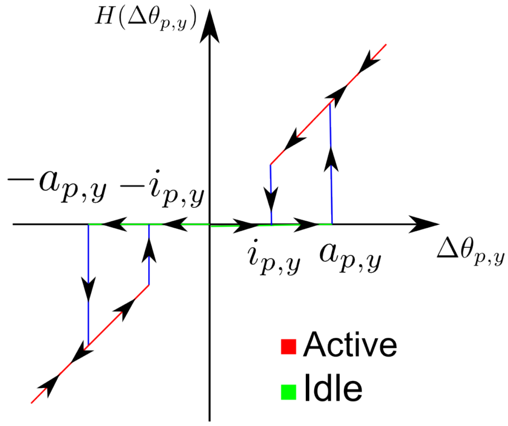

Figure 2.10: Hysteresis transfer function. In red is represented the active mode and in green the idle mode for each angle pan and tilt.

The switch between the two modes happen when:

- $-i_\alpha < \Delta\Theta_\alpha < i_\alpha$ the controller enters idle mode,
- $-a_\alpha > \Delta\Theta_\alpha > a_\alpha$ the controller enters active mode,

with $\alpha \in \{p, y\}$.

In the active mode the block output $H(\Delta\Theta_{py})$, is equal to the input $\Delta\Theta_{py}$. In this mode the controller will reduce $|\Delta\Theta_\alpha|$ to $i\alpha$ by moving the pan and tilt unit to the same orientation has the HMD. In the idle mode, the output $H(\Delta\Theta_{py})$ is zero. This mode prevents small head movements induce a motor movement and only the VPTR block will correct the stereo image perspective. The boundary values a_α and i_α are adjustable per angle to provide a smooth transition between the two modes.

The next block is the proportional block, K. This block converts the error $\Delta\Theta_{py}$ to motors motors velocity, v

$$v_{py} = K_{py}\,\Delta\Theta_{py}. \tag{2.17}$$

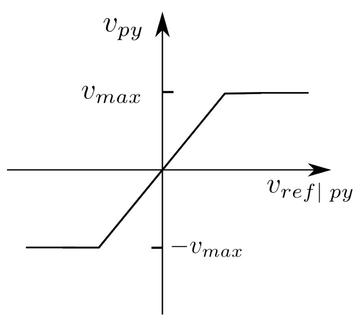

Figure 2.11: Saturation block transfer function.

The next block is the saturation block. This block limits v_{ref} in py to the motor maximum speed, v_{max}.The purpose of this block is to assure that the velocity calculated on the controller is never superior to the motors maximum speed. This way one can assure the position sent to the motors by the controller will be achieved. The transfer function of this block is presented in figure 2.11.

The last block is an integrator. The integrator converts v_{py} into the position, Θ_{PTU} to be sent to the motors. The time response is given by

$$\Theta_{PTU}(t) = \Theta_0 + \int_0^t v_{py}(t) \tag{2.18}$$

and in discrete form

$$\Theta_{PTU}(t) = \Theta_{PTU\ t_{-1}} + v_{py}\ \Delta(t - t_{-1}) \tag{2.19}$$

Since the PTU do not provide any feedback with the current position or velocity Θ_{PTU} is also used as the motors position to close the controller loop. This can only be done because of the saturation block, that assures that the velocity in the controller is never superior to the motors maximum speed. With this restrain one can assume that the motors always achieve the desired position calculated through the controller.

Chapter 3

Experiments and Results

The goal of this experiment is to measure the Situational Awareness (SA) of an operator while teleoperating a robot with different setups. SA is not a measurable entity but can be inferred from experience results. This chapter follows with a section with the experience protocol and another section to present the results.

3.1 User study

The objective of the experience was to test three visualization systems with RAPOSA. Each volunteer was required to teleoperate RAPOSA three times to complete a search mission, each with one of the following systems:

- GUI – Using a GUI in a computer screen,
- HMD – Wearing a Head mounted display with an attitude sensor providing a direct control of the stereo camera orientation,
- HMD + VPTR – The control method developed on this thesis that blend the stereo camera orientation control with the virtual pan,tilt and roll method.

Each mission lasted 5 minutes and during that time the volunteer task was to identify the max number of victims possible. A victim was considered successfully identified when the volunteer could say the number associated to that victim. Between missions and to ensure that the volunteer make a correct identification, the number associated to each victim changed. The scenario was inspired in NIST RoboCup scenario[1] . The scenario was based on the yellow arena and was designed to provide the same visual challenge. To ensure unbiased result of the scenario learning from one test to the other, the visualization systems were tested in different orders, two for each combination, which performed a total of 12 volunteers and 36 missions. For each mission the only knowledge passed to the volunteer is an explanation on the system he was going

[1] `http://www.isd.mel.nist.gov/projects/USAR/arenas.htm`

to use. The volunteer did not had any *a priori* knowledge of the scenario or map, except the one he acquired while performing the previous mission. Between each mission volunteers where asked to reply to NASA-TLX questionnaire, a standard question used by NASA to measure task payload. In the end each volunteer is required to answer a questionnaire. An example of the questionnaire can be found in the appendix.

Figure 3.1: User study Scenario: image from the scenario where the user study was performed.

3.2 Results and discussions

The questionnaires results where analyzed by performing the mean result and a single-sided paired test to obtain the statistical significance for each of the five criteria analyzed: Image Quality/Resolution, 3D Perception, Perception of the Environment around the Robot, Ease to Identify Victims and Ease in Control the Robot. The volunteers evaluated those criteria in a 1 to 10 scale, where 1 means bad and 10 excellent. The results are shown in table 3.1.

Table 3.1: Mean results of the criteria evaluation by the volunteers after testing the 3 systems.

Criteria	GUI	HMD	HMD+VPTR
Image Quality/Resolution	8.42	4.25	6.17
3D Perception	1.58	4.67	7.42
Perception of the Environment Around the Robot	4.75	4.83	6.67
Ease Identifying Victim	7.33	4.42	6.75
Ease in Control the Robot	6.83	6.25	7.50

Image Quality/Resolution

In terms of image quality the GUI had a superior advantage due a superior image resolution since the images did not suffer any manipulation and were directly showed on the computer screen.There was a statistically significant difference between systems as determined by one-way ANOVA F(2,33) = 25.023, p =0. The Bonferroni *post-hoc* test revealed that GUI had the best result (U=8.42±0.31, $p < 0.01$). With the GUI system, the volunteers did not have to get so close to the victims in order to identify them has they had with the other two systems. Although the HMD had a larger field of view, when comparing with the HMD+VPTR, the last had a better score (HMD+VPTR U=6.17±1.33 , $p < 0.01$). This result suggests that the stable image and the smooth movement performed by the camera with the HMD+VPTR enables a better perception of the environment overcoming the lower image quality.

3D Perception

In this category HMD+VPTR outperformed over the other two systems. The one-way Anova was statistically significance with F(2,33) = 42.267, p =0. The Bonferroni *post-hoc* test revealed that the HMD+VPTR system (U=7.42±1.24 , $p < 0.01$) performed much better than HMD (U=4.6±2.1 , $p < 0.01$) . This result is coherent with the expected. The GUI does not provide any 3D experience and the operators could only count on their common sense to perceive how close they were to obstacles. Since the HMD+VPTR images were more stable and did not flick, it was more comfortable to use than the HMD.

Perception of the Environment Around the Robot

Perception of the environment is not a measurable quantity and the results interpretation can be misleading. To evaluate this criteria the two main aspects to take in account are the image quality and how easy it is to explore and perceive the environment around the robot. The HMD and HMD+VPTR systems have the advantage on the environment exploration due to the camera control with the head attitude, making the environment exploration more intuitive. The one-way Anova was statistically inconclusive but the t-test was performed. When comparing the HMD with the GUI the GUI got a better score. This can be the result of the superior image quality of the GUI. But when comparing GUI with HMD+VPTR, the HMD+VPTR have a better result(HMD+VPTR U=6.67 , $p < 0.01$). This result shows that even with a worst image quality and a narrow field of view, the smoother and intuitive camera control proved to be good combination to improve the perception of the environment around the robot for the volunteers.

Ease Identifying Victims

Identifying victims was not a simple task. Even when the volunteer was repeating the test he needed to get close to the victim to be able to see the associated number. The one-way Anova was statistically significance with $F(2, 33) = 11.526, p = 0$. In this task the GUI system was in advantage because of the superior image quality. But following the previous criteria, the HMD+VPTR outperformed both the other systems (HMD+VPTR U=6.75±1.5 , $p < 0.01$). This result come from the safe the operator feel when controlling the robot near obstacles with the HMD+VPTR system. Although they had to get closer to the victims they could perform it in a secure way surpassing the GUI superior image quality.

Ease in Control the Robot

The last criteria relates to how secure the volunteers feel in moving the robot around the scenario.The one-way Anova was statistically inconclusive but the t-test was performed. Once again HMD+VPTR outperformed the other two systems (HMD+VPTR U=7.5 , $p < 0.01$). Although the image quality shows its importance in this criteria when comparing the HMD with the GUI, where the GUI got a better score, the volunteers evaluated the HMD+VPTR as a better system to control the robot.

Nasa-TLX Results

The results from this questionnaire give information about the systems in how stressful and hard was to accomplish the tasks. Due to an error when answering the performance question by some of the volunteers this question did not provide any real information and was not taken in this analysis. In NASA-TLX each factor was rated from 1 to 20 being 20 very bad and 1 very good. The results are shown in the table 3.2

Table 3.2: Mean results of the NASA-TLX questionaire for the three systems.

Factor	GUI	HMD	HMD+VPTR
Mental Demand	6.42	8.33	6.00
Physical Demand	2.83	7.5	7.20
Temporal Demand	8.42	10.83	9.9
Effort	5.42	8.92	7.8
Frustration	3.75	5.92	5.2

The GUI system was considered to require less physical demand, effort and was less frustrating to teleoperate the robot with. The physical demand is intuitive because the user did not have to wear an HMD and to move his head while performing with this system. With the better quality image, the user could identify the victims without getting too much closer which result in HMD to be a much less effort and frustrating system. The HMD+VPTR outperformed the HMD in all the factors showing that this system really improves and eases the teleoperation task. When comparing with the mental demand factor, the HMD+VPTR got a better result than the GUI system.

Likert Scale Sentences Classification

Figure 3.2 shows the result of the Likert scale questions where the volunteers where asked to compare the three systems by rating a sentence through Likert scale, classifying each sentence from ”Strongly Disagree” to ”Strongly Agree”. These rate were then transformed to a rate from one to five respectively and the result plotted on a boxplot graphic. The sentences were numbered from Q4.1 to Q4.4:

- Q4.1 – The usage of HMD helps me better understand the position of RAPOSA,

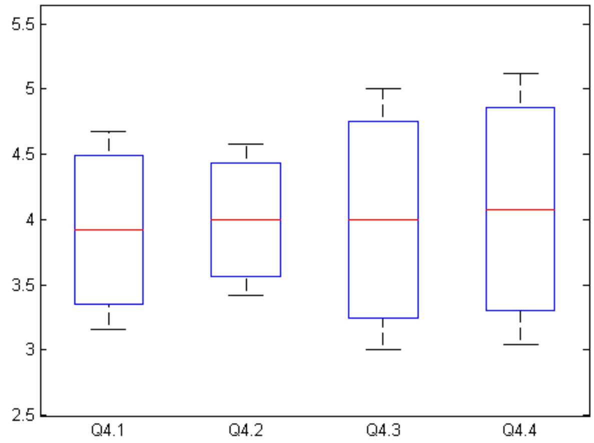

Figure 3.2: Likert scale chart.

- Q4.2 - The usage of HMD helps me better understand the structure of the enviorment,
- Q4.3 - The robot teleoperation is easier using HMD+VPTR over HMD,
- Q4.4 - The visual perception of the enviorment is more intuitive with HMD+VPTR over HMD.

The first two sentences try to understand if the volunteers opinion if wearing a HMD was helpful to teleoperate RAPOSA. The last two sentences were to evaluate the HMD versus the HMD+VPTR systems. Following the results previously obtained, the operators classified the HMD helpful to teleoperate RAPOSA (Q4.1: U=3.92, σ = 0.76; Q4.2: U=4, σ = 0.58). Although the last two sentences had a bigger standard deviation (Q4.3: U=4, σ = 1; Q4.4: U=4.08, σ = 1.04), the mean results were superior, classifying the HMD+VPTR as a better system to teleoperate RAPOSA than HMD.

Volunteers feedback

After performing each mission and through the hole evaluation process the volunteers were providing feedback about the systems used. About the GUI system the volunteers referred the that the image quality really helped with the task allowing to easily identify the victims. About the decoupling of the camera control it was referred that "*I can drive the robot and move the camera at the same time*". "*If the HMD+VPTR had the same image quality has the GUI I would prefer the HMD+VPTR.*" and similar sentences were stated by several volunteers. Some people said that to drive the robot they would prefer the HMD+VPTR but to identify the victims the GUI was a better system. This shows the importance of the image quality while teleoperating a robot. Two of the volunteers stated that the roll provided by the HMD+VPTR make the teleoperation more intuitive, "*With roll this system gets so intuitive!*". These feedback provided by the volunteers reinforce the results previously presented.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

This work presents the development of a system to control a stereo camera orientation mounted on a pan and tilt system through the orientation provided by an HMD attitude sensor. It also presents a virtual pan, tilt and roll method that simulates the camera movement to HMD user. The proposed controller blends both physical and virtual pan and tilt enabling a faster response for the operator head movements and mitigating the pan and tilt system physical limitations. A case study was performed to test the system. The case study showed that the image quality is a very important aspect to enhance a user perception of the environment. During the whole case study the volunteers felt more comfortable exploring the scenario with the development system that showed to be a more intuitive way to control the robot due to the decoupling of the camera control. The virtual pan tilt and roll method showed to enhance the user perception by reducing the pan and tilt system limitations and enabling a faster response to the user movements. Due to the simple design and since the controller architecture does not depend on the system components the virtual pan, tilt and roll and the controller can be applied in similar tasks to obtain similar results.

4.2 Future Work

It would be interesting to repeat the a similar experience with the system proposed in this thesis but with an HMD with a superior image quality and attitude sensor. It would also be very interesting to introduce Augmented Reality in the system by projecting in the HMD screens the path that the robot will follow if the operator keep forward. It would also be interesting to test a multi camera teleoperation system: one to teleoperate the robot while exploring a scenario and one to see closer objects. This last camera could be mounted in a controlled arm and the operator could choose between the two cameras.

Bibliography

[1] Henrique Martins and Rodrigo Ventura. Immersive 3-d teleoperation of a search and rescue robot using a head-mounted display. 2009.

[2] Jussi Suomela. Tele-presence aided teleoperation of semi-autonomous work vehicles. Technical report, 2001.

[3] Elizabeth S. Redden Linda R. Elliott, Chris Jansen and Rodger A. Pettitt. Robotic telepresence: Perception, performance, and user experience. 2012.

[4] J. Larry Pezzaniti, Richard Edmondson, Justin Vaden, Brian Hyatt, David B. Chenault, Joseph L. Tchon, Tracy J. Barnidge, and Brad Pettijohn. Flat panel 3d display for unmanned ground vehicles. pages 73320N–73320N–12, 2009.

[5] M. Eng. Massachusetts Institute of Technology Sinn, Richard. Virtual pan-tilt-zoom for a wide-area-video surveillance system, 2008.

[6] Mattias Seeman, Mathias Broxvall, and Ro Saffiotti. Virtual 360 panorama for remote inspection, 2007.

[7] Michael Jenkin, James Elder, and Greg Pintilie. Loosely-coupled telepresence through the panoramic image server, 1998.

[8] Shenchang Eric Chen. Quicktime vr: an image-based approach to virtual environment navigation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '95, pages 29–38, 1995.

[9] Andrea Fusiello, Emanuele Trucco, Alessandro Verri, and Ro Verri. A compact algorithm for rectification of stereo pairs, 1999.

[10] Frank Steinicke, Gerd Bruder, Scott Kuhl, Pete Willemsen, Markus Lappe, and Klaus Hinrichs. Natural perspective projections for head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 17(7), July 2011.

[11] Richard Edmondson, J. Larry Pezzaniti, Justin Vaden, Brian Hyatt, James Morris, David Chenault, Andrew Bodenhamer, Bradley Pettijohn, Joe Tchon, Tracy Barnidge, Seth Kaufman, David Kingston, and Scott Newell. 3d display for enhanced tele-operation and other applications. pages 76901D–76901D–10, 2010.

[12] Ludek Zalud. Argos system for heterogeneous mobile robot teleoperation. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2006, October 9-15, 2006, Beijing, China*, pages 211–216. IEEE, 2006.

[13] Richard Edmondson, Todd Aycock, and David Chenault. Auto-converging stereo cameras for 3d robotic tele-operation. pages 838414–838414–7, 2012.

[14] Terrence Fong and Charles Thorpe. Vehicle teleoperation interfaces. *Auton. Robots*, pages 9–18, 2001.

[15] Charles Loop and Zhengyou Zhang. Computing rectifying homographies for stereo vision, 1999.

[16] M Axholt, M A. Skoglund, Stephen D. O'Connell, Matthew D. Cooper, Stephen R. Ellis, and Anders Ynnerman. Parameter estimation variance of the single point active alignment method in optical see-through head mounted display calibration. In *Proceedings of the 2011 IEEE Virtual Reality Conference*, pages 27–34, 2011.

[17] W. N Kama and R. C. Dumars. Remote viewing: A comparison of the direct viewing, 2d and 3d television.

[18] Jean Scholtz and Jeff Young. Evaluation of human-robot interaction awareness in search and rescue. In *In Proceedings of the 2004 International Conference on Robotics and Automation*, 2004.

[19] Carlos Candido, Pedro Santana, Luís Correia, and José Barata. Shared control of a pan-tilt camera on an all-terrain mobile robot. In *ETFA*, pages 177–183. IEEE, 2008.

[20] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty: A review. *Int. J. Comput. Vision*, 27:161–195, 1998.

[21] S. Arca, E. Casiraghi, and G. Lombardi. *Corner localization in chessboards for camera calibration*. IADAT, 2005.

[22] Masayuki Kanbara, Takashi Okuma, Haruo Takemura, and Naokazu Yokoya. A stereoscopic video see-through augmented reality system based on real-time vision-based registration. In *System Based on Real-time Vision-based Registration," Proc. IEEE Virtual Reality 2000*, pages 255–262, 2000.

[23] A. Bernardino and J. Santos-Victor. Binocular tracking: integrating perception and control. *Robotics and Automation, IEEE Transactions on*, 15(6):1080–1094, 1999.