2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Design and Evaluation of a Telepresence Vision System for Manipulation Tasks

Koji Shiratsuchi, Kohei Kawata, Emmanuel Vander Poorten and Yasuyoshi Yokokohji

Abstract— **This paper describes the design of a new telepresence vision system developed to realize a higher immersive feeling for telemanipulation tasks. A new measure of 'permissible visual errors' was defined. Making use of this new measure, a minimal vision system is designed, containing only the strictly necessary DOF's (degrees of freedom) while keeping the vision errors below an experimentally obtained set of permissible errors. The result is a 4DOF camera system, containing two rotational joints (pan-tilt) and two prismatic joints (horizontal plane.) An evaluation of the system was done through a telemanipulation task using a unified hand/arm teleoperation testbed. It was found that during use of the vision system almost all visual errors remained within the permissible errors. The proposed vision system design framework suggests to break away from simply (and often blindly) mimicking human appearance. Although we only dealt with the vision system design, the same concept could be used when designing any other part of a robot.**

I. INTRODUCTION

Teleoperated robots have been used in hazardous environments, such as nuclear power plants, space, the bottom of sea, where autonomous robots cannot perform the required tasks due to the complexity of these highly unstructured environments. Ideally, a teleoperation system should provide an immersive feeling to the operator by providing several modalities (vision, sound, kinethesia,...) as if the operators were directly performing the task in the remote environment.

According to this 'telepresence' concept, slave robots should have the same configuration as human operators. So in a sense, humanoids can be seen as ideal slave robots. However, even for such robots, operation can be tedious and tiring[1]. Also, building a slave with the so many DOF is costly and complex. For example, HRP-1, a teleoperated humanoid robot[2], has 30 DOF in total, still much less than humans have. On the other hand, robots do not have to perform all movements that humans do. If a target task is strictly given, it becomes possible to extract the essential number of DOF and to simplify the robotic system.

Several approaches can be taken to realize a more immersive feeling in a tele-existence system: improvements to arm designs, adaptations to controllers or addition of extra audio, visual or tactile information[3], [4]. Especially the vision system plays a crucial role as it delivers a vast amount of information from the remote environment. Lots of efforts has been done already to improve vision systems. The use of a head mount display (HMD) to display the remote scene was found to be superior to display by conventional cathode ray tube monitor[5]. Control methods of the camera system to maximize information[6], or to reduce mental rotations and translation (mental workload) were proposed[7]. Vergence control of a slave camera system was introduced in [8] to improve the depth perception of the remote environment.

Koji Shiratsuchi, Kohei Kawata, Emmanuel Vander Poorten and Yasuyoshi Yokokohji are with the Department of Mechanical Engineering and Science, Kyoto University, Mechatronics Laboratory. yokokoji@mech.kyoto-u.ac.jp

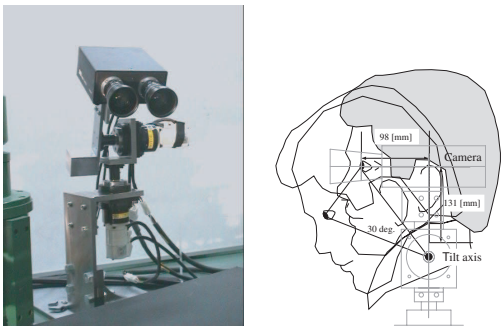


Fig. 1. Camera head design for the original system

At present, a big majority of 2-DOF (pan-tilt) and 3-DOF (pan-tilt-roll) slave camera systems are built without much consideration about needs, or optimal configuration. Although human's capability to unconsciously adapt to the given camera head system might initially cover for potential problems, these systems will make the operator feel sick and lead to inefficient operations in the end.

In this paper, we introduce a new approach to design vision systems by carefully observing the head motions of the human operator. These observations taught us two important facts. The first is that people tend to translate their head wider than usual, when they are working with a narrow field of view (in our system approximately 48°). The second fact is that people are not sensitive to small errors in position and orientation and practically can't perceive these. We call these errors 'permissible visual errors'.

We propose to design vision systems by comparing the expected visual errors with the 'permissible visual errors'. An ideal system is one that keeps the visual errors within the permissible, and only uses a minimal number of DOF's to do this. Such system will show a higher immersive feeling and high user-friendliness, improvements in task performance and accuracy will be a logical consequence.

In the following, section II introduces the unified telemanipulation testbed. In section III the new measure of permissible errors is discussed. Section IV describes the optimization of a system's structure accoring the new measure. Practical evaluation on a teleoperation experiment is described in V. Finally, conclusions are drawn and directions for future work are sketched in VI.

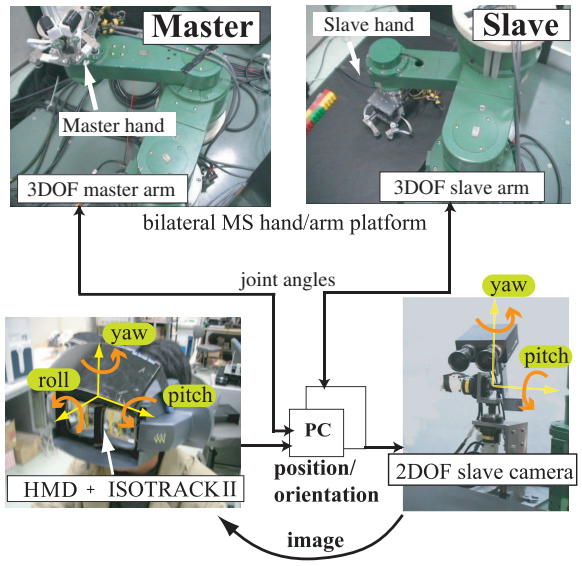


Fig. 2. unified teleoperation testbed

II. UNIFIED HAND/ARM MASTER-SLAVE TELEOPERATION TESTBED

A. ORIGINAL SYSTEM SETUP

Fig.2 shows the teleoperation testbed, originally developed by Yokokohji et al.[9] and used in this study for evaluating the effectiveness of the proposed approach. The testbed is composed of a unified hand/arm master-slave system and a master-slave camera system. The hand/arm system consists of a pair of 3-DOF SCARA-type direct-drive arms extended with two 4-DOF hands. All joints are actuated and bilateral control is implemented. An HMD (STV-E, SHIMADZU Co.) is used at the master side and a stereo camera is installed beside the slave arm to visualize the task environment. A 6DOF magnetic sensor, ISOTRACK II(Polhemus Inc.) is used to detect the position and orientation in roll(ϕ), pitch(θ) and yaw(ψ), at a sampling rate of 10Hz.

The camera (KS55 Kastam Co., 45 deg field of view (FOV), 1.8kg) is mounted on a 2-DOF camera head, which was designed based on the neck-head configuration of the human. The system captures yaw and pitch-DOF, being most frequently used by human's. The location of the joints and link lengths were also decided in correspondence to the human as shown in Fig.1.

Actually, it turned out that this blind mimicking of the human head led to insufficient quality of the obtained images. A lack of immersivity was found as performance of tasks when using camera-system and HMD was much worse than for the case were direct sight was used. The subsections below describe some reasons for this lack of immersivity.

B. EFFECT OF REDUCED FIELD OF VIEW

A major explanation for the loss of immersivity can be found in the reduction of the operator's FOV. Although the structure of the original camera head corresponds well to that of a human, the HMD reduces the FOV strongly. Following experiment was done to measure the effect of the reduced FOV on a human operator's head motions.

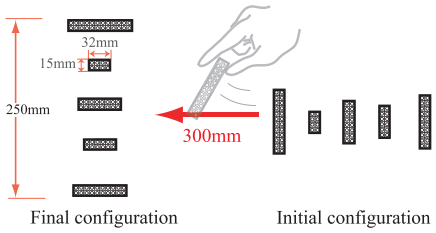


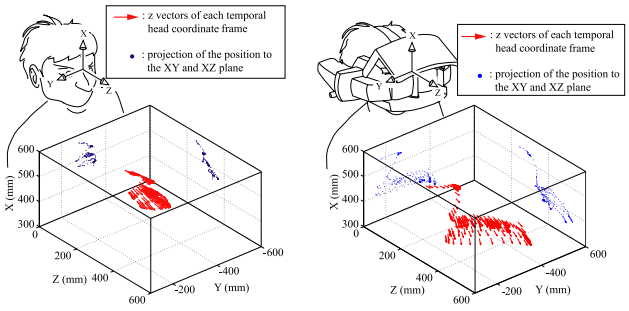
Fig. 3. Rearranging the blocks

A benchmark teleoperation task, utilizing LEGO blocks[9] was designed. Subjects were asked to rearrange a number of LEGO blocks from an initial position to a predefined target position (Fig.3). First, the experiment was done with normal FOV (direct sight) and then with restricted FOV. For the latter, the same HMD as in normal teleoperation was worn, but now used as a see-through, by opening the optical shutter (direct sight + restricted FOV) as shown in Fig.4. The user's head movements were measured by the ISOTRACK II. Fig. 4 depicts a typical scenario of the users head movements.

Four males in the 20s participated in the experiment, performing each task 5 times. The experimental results are summarized in Table I, where directions X, Y, Z correspond to those indicated in Fig.4. An Analysis of variance (ANOVA) was done to verify the confidence of the results. It was found that especially movements in the Z direction increase (α-value smaller than 0.01), whereas increase of movements in other directions is less outspoken and restricted to a limited number of participators.

C. EFFECT OF SIMPLIFIED STRUCTURE

Another explanation for the loss of immersivity can be found in the simplifications done when designing the camera system. The roll-DOF was removed and the location of the tilt axis was put at the base of the human's neck. In reality a neck is a chain of parallel joints, so when approximating this by one rotational joint, its location should be optimized to cover as close as possible the movements of a real neck.



(a) Normal field-of-view (b) Narrowed field-of-view

Fig. 4. Measured head movement

TABLE I
ANALYSIS OF VARIANCE

(mm)	Subject 1	Subject 2	Subject 3	Subject 4
X	0.73	3.93	2.36	70.69**
Y	3.61	12.57**	0.00	6.21*
Z	54.56**	25.32**	15.14**	45.82**

*: Rejection region in $\alpha = 0.05$, **:Rejection region in $\alpha = 0.01$

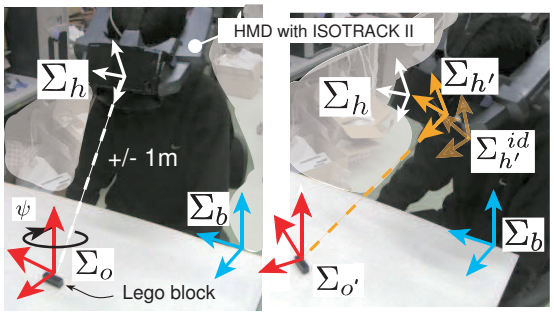


Fig. 5. Measuring permissible visual errors experimentally

To conclude, it is clear that intuitive designs of camera systems are insufficient. An objective measure of the visualization quality is needed, allowing designers to determine the minimal DOFs to realize a good visualization for a specified set of tasks or optimize various link parameters.

III. PERMISSIBLE ERRORS

In this section a new measure to determine the visualization quality is defined (subsection III-A). Subsection III-B describes a practical calculation method of the new measure and examples are given in III-C.

A. DEFINITION OF PERMISSIBLE ERRORS

Definition 1: The permissible errors e_{prm} of a certain property of a certain task are deviations from this property that remain unnoticed by a majority of human observers. These from a set[1]

$$e_{prm} \in \left[E[\Delta e] \pm n\sigma(\Delta e)\right], \tag{1}$$

where Δe is the size of an unnoticed error and $E[\Delta e]$ and $\sigma(\Delta e)$ are the bias[2] and standard deviation of it's distribution. The parameter n is a coefficient that can be selected to specify the size of the group of observers for which the property variation goes unnoticed[3].

Let us now clarify this definition by discussing a concrete example of a visualization task. Suppose that a human subject is observing a target object as shown in Fig.5(a) (called the 'nominal view'). The permissible error on the pose of the observed object is then a measure of the size of variation on the object's pose that is not perceived by the user. Following method can be used to measure this:" after changing the pose of the target object, we ask the subject to move his head until he thinks to see the target object with the same appearance as before (Fig.5b)." This second situation is called the 'secondary view'. Now, by measuring the head and target object pose in both views, the object pose with respect to the head coordinate frame $\mathbf{\Sigma}_h$ is found as ${}^h\boldsymbol{T}_o$ and the relative pose with respect to $\mathbf{\Sigma}_{h'}$ is ${}^{h'}\boldsymbol{T}_{o'}$. Where subscripts o and h stand for object and head respectively. A prime-mark such as in o' and h' is added, whenever the value in the secondary view is meant. For an ideal relocation of the human's head, ${}^{h'}\boldsymbol{T}_{o'}$ should be identical to ${}^h\boldsymbol{T}_o$ and the object's image will be the same. In reality, human's vision capability is limited and the relocation is not perfect. The permissible error for this visualisation task can then be calculated by applying (1), where Δe is the difference between the visualization of the object in ideal ${}^{h'}\boldsymbol{T}_{o'}^{id} = {}^h\boldsymbol{T}_o$ and actual pose ${}^{h'}\boldsymbol{T}_{o'}$.

[1]Note that a permissible error is thus a task and property specific criterion. Design of systems based on this criterion asks for an identification of the relevant tasks and properties to be performed and an experimentally measurement of the permissible errors at beforehand.

[2]In principle there should be no bias at all. However if it does exist, it is important to investigate the reason for its presence. In case of measurement error, the bias should be neglected, however in case when human or physical factors can be identified as being the cause, the bias should be incorporated.

[3]For $n = 0$ and $E[\Delta e] = 0$ the system should be perfect as no error is allowed. When increasing n the number of people who don't notice the difference decreases, e.g. assuming a Gaussian distribution of the error for $n = 0.385$ this is 70%, for $n = 0.675$ this is 50%.

B. DERIVATION

Next, calculation methods for permissible errors are derived. First, position errors are treated, an explanation for rotational errors follows. The method follows above example for poses in nominal and secondary views.

1) Permissible Position Errors: In the nominal view, the object pose with respect to $\mathbf{\Sigma}_h$, ${}^h\boldsymbol{T}_o$, can be calculated as

$${}^h\boldsymbol{T}_o = ({}^b\boldsymbol{T}_h)^{-1}\ ({}^b\boldsymbol{T}_o). \tag{2}$$

Where ${}^b\boldsymbol{T}_h$ and ${}^b\boldsymbol{T}_o$ are transformation matrices representing the head and object's pose, expressed in an invariant base frame $\mathbf{\Sigma}_b$. The former can be measured by for example a magnetic tracker, the latter follows from the task setup. In a similar way ${}^{h'}\boldsymbol{T}_{o'}$ can be determined as

$${}^{h'}\boldsymbol{T}_{o'} = ({}^b\boldsymbol{T}_{h'})^{-1}\ ({}^b\boldsymbol{T}_o\ {}^o\boldsymbol{T}_{o'}), \tag{3}$$

with ${}^o\boldsymbol{T}_{o'}$ the transformation matrix representing the object pose change from nominal to secondary view.

Next, we derive numerically what the effect is of a deviation of the actual pose ${}^{h'}\boldsymbol{T}_{o'}$ from the ideal head pose ${}^{h'}\boldsymbol{T}_{o'}^{id}$ on the object's 'image' as seen by the user. Fig.6 supports the explanation. A similar figure can be drawn for the secondary view by adding primes' to all appropriate parameters. The origin of $\mathbf{\Sigma}_o$ is located at the center of the object and $\mathbf{\Sigma}_h$ is attached just between the two eyes. An 'object plane' and a 'virtual image plane' (VIP) are defined as planes parallel to the x-y plane of $\mathbf{\Sigma}_h$. The object plane goes trough the origin of $\mathbf{\Sigma}_o$ and the distance to $\mathbf{\Sigma}_h$ is defined as D_h (or D'_h in secondary view), whereas the distance between VIP and $\mathbf{\Sigma}_h$ is set, without losing generality, equal to 1. The origins of the planes are located at the intersection with the axis $\boldsymbol{z}_h$. ΔD is defined as the relative difference

$$\Delta D = \frac{1}{D_h}(D_h - D'_h). \tag{4}$$

X_h, Y_h (and $X_{h'}, Y_{h'}$) are the coordinates of the target object in the object plane. Their visualization errors (in VIP) are

$$\Delta X = \frac{X_h}{D_h} - \frac{X'_h}{D'_h} \quad \text{and} \quad \Delta Y = \frac{Y_h}{D_h} - \frac{Y'_h}{D'_h}. \tag{5}$$

Actually, X_h, Y_h and D_h are nothing more than the coordinates of ${}^h\boldsymbol{p}_o = (X_h\ Y_h\ D_h)^T$ and similarly in the secondary

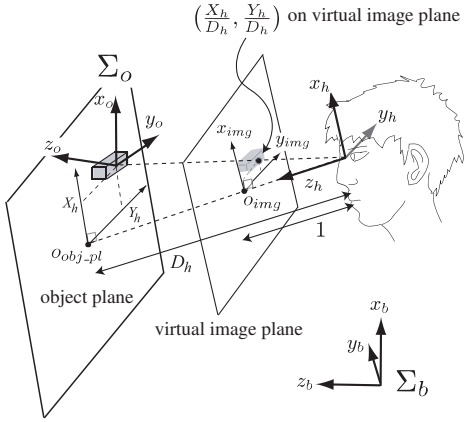


Fig. 6. Permissible errors and Virtual Image Plane

view ${}^{h'}\boldsymbol{p}_{o'}=(X'_h\ Y'_h\ D'_h)^T$,which can be found from

$$ {}^{h}\boldsymbol{T}_o = \begin{bmatrix} {}^{h}\boldsymbol{R}_o & {}^{h}\boldsymbol{p}_o \\ \boldsymbol{0} & 1 \end{bmatrix} \quad {}^{h'}\boldsymbol{T}_{o'} = \begin{bmatrix} {}^{h'}\boldsymbol{R}_{o'} & {}^{h'}\boldsymbol{p}_{o'} \\ \boldsymbol{0} & 1 \end{bmatrix}. \quad (6) $$

2) Permissible rotational errors: The 2-axis rotation method[10] is used to describe the orientation errors. In contrast to Euler angles, which represent the object orientation by a series of three rotations $(\phi_{eu}, \theta_{eu}, \psi_{eu})$ around Z-,Y- and Z-axis, the 2-axis rotation method allows us to describe the orientation with only two steps of rotation. Each rotation angle can be obtained from the Euler angles as $(\theta_{eu}, \phi_{eu} + \psi_{eu})$. A further advantage of the 2-axis rotation method is that, under the assumption that θ_{eu} is small, the two angles correspond to a rotation about an axis in and a rotation about an axis perpendicular to the image plane. ${}^{h}\boldsymbol{R}_o$ and ${}^{h'}\boldsymbol{R}_{o'}$ are known from (6). The error of these two rotation matrices can be described by ${}^{h}\boldsymbol{R}_o\ ({}^{h'}\boldsymbol{R}_{o'})^{-1}$. For corresponding Euler angle errors, $\Delta\phi, \Delta\theta, \Delta\psi$, follows

$$ {}^{h}\boldsymbol{R}_o\ ({}^{h'}\boldsymbol{R}_{o'})^{-1} = \begin{bmatrix} C_{\Delta\phi}C_{\Delta\theta}C_{\Delta\psi} - S_{\Delta\phi}S_{\Delta\psi} & -C_{\Delta\phi}C_{\Delta\theta}S_{\Delta\psi} - S_{\Delta\phi}C_{\Delta\psi} & C_{\Delta\phi}S_{\Delta\theta} \\ S_{\Delta\phi}C_{\Delta\theta}C_{\Delta\psi} + C_{\Delta\phi}S_{\Delta\psi} & -S_{\Delta\phi}C_{\Delta\theta}S_{\Delta\psi} + C_{\Delta\phi}C_{\Delta\psi} & S_{\Delta\phi}S_{\Delta\theta} \\ -S_{\Delta\theta}C_{\Delta\psi} & S_{\Delta\theta}S_{\Delta\psi} & C_{\Delta\theta} \end{bmatrix}, \quad (7) $$

where $C_{\Delta\phi}$ and $S_{\Delta\theta}$ are $\cos(\Delta\phi)$ and $\sin(\Delta\theta)$, respectively, and so on. Once the Euler angles of the orientation error are found, the two error angles are $\Delta\theta$ and $(\Delta\phi + \Delta\psi)$.

By measuring the head pose in the secondary view many times, statistical data for ΔD, ΔX, ΔY, $\Delta\theta$ and $\Delta\phi + \Delta\psi$ is obtained. The permissible errors for ΔD_{prm}, ΔX_{prm}, ΔY_{prm}, $\Delta\theta_{prm}$ and $(\Delta\phi + \Delta\psi)_{prm}$ can then be obtained by applying (1) for a specified value of n. When designing vision systems, all the visual errors should be kept within the permissible errors, which can be expressed as:

$$ \Delta D < \Delta D_{prm},\ \Delta X < \Delta X_{prm},\ \Delta Y < \Delta Y_{prm}, $$
$$ \Delta\theta < \Delta\theta_{prm}, \quad (\Delta\phi + \Delta\psi) < (\Delta\phi + \Delta\psi)_{prm}. \quad (8) $$

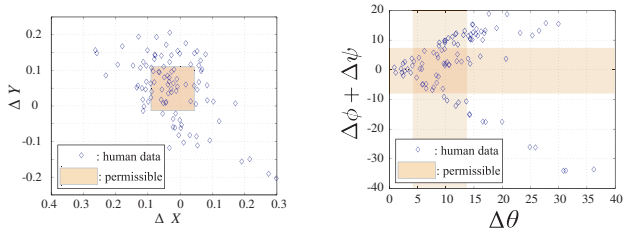
C. Experiment

The permissible errors for the task of positioning LEGO blocks (Fig.3) were determined. The subjects were asked to observe a block carefully (nominal view), and then reposition themselves after the block was displaced, until they claimed that the appearance of the block was identical to the initial view. The subject's head position and orientation

TABLE II

PERMISSIBLE ERRORS FOR 5 TASKS

	ΔX	ΔY	ΔD	$\Delta\theta_{eu}$	$\Delta\phi_{eu} + \Delta\psi_{eu}$
Mean	-0.024	0.050	0.095	12.27	1.88
Std.Dev.	0.096	0.089	0.0736	7.188	11.41
Δe_{prm}	$\pm$0.065	$\pm$0.06	$\pm$0.05	9.70	$\pm$7.69



(a) position errors (b) rotation errors

Fig. 7. Errors on the virtual image plane

was measured with the ISOTRACK II. Four males in the 20s participated in the experiment. They were asked to visually cancel the following 5 types of block displacements:

$$ \boldsymbol{R}({}^{b}X_o, 30°),\quad \boldsymbol{R}({}^{b}X_o, -30°),\quad \boldsymbol{R}({}^{b}Y_o, 30°), $$
$$ \boldsymbol{R}({}^{b}Y_o, 30°)\boldsymbol{R}({}^{b}X_o, 30°),\ \boldsymbol{R}({}^{b}Y_o, 30°)\boldsymbol{R}({}^{b}X_o, -30°) \quad (9) $$

The tasks were executed 5 times each, in total 100 data points were gathered. The initial position of the LEGO block was at $\approx 1m$ from $\boldsymbol{\Sigma}_h$. The permissible errors were calculated based on (1). A value of $n = 0.675$ was taken. In such a case at least 50% of the subjects won't notice any error. Note that this does not necessarily mean that the other 50% is dissatisfied. Fig. 7 and Table II summarize the experimental results. The rotational errors are also described with the 2-axis method. $\Delta\theta$ is always taken $\in [0°, 180°]$. Ideally its distribution must show a peak at 0, decreasing over the positive domain. The bias appearing in our experiments caused a shift of the peak inside the positive domain. To avoid overestimating the permissible region, we ignored the bias and took a range around the peak to describe the permissible error for $\Delta\theta$.

IV. SLAVE CAMERA DESIGN BASED ON THE PERMISSIBLE ERRORS

In this section, the design of a new camera system, based on the obtained estimates for the permissible errors, is described. The experiment showed that especially translational movements should be covered better by the camera system. Rather than adding extra rotational DOFs, we chose to use extra translation joints. In this way the visualization problem can be decoupled in a position and orientational part. Also from a control point of view this simplifies the problem substantially. First, the design of the orientational (IV-A) and then the translational part (IV-B) is described.

A. OPTIMIZATION OF ROTATIONAL PART

From Table II it is clear that one rotational DOF is insufficient to keep the visualization errors below the permissible errors. We look therefore at 2-or-more-DOF camera systems. First, investigation of 2-DOF system was done. Because, if a 2-DOF system can be found that keeps the visualization

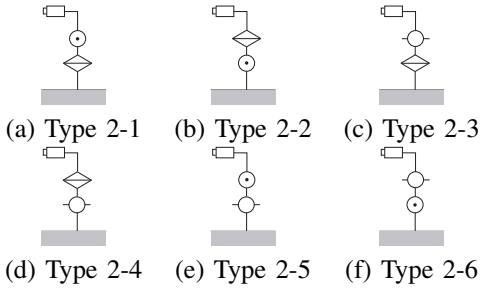


Fig. 8. Candidates of 2-DOF slave camera system

errors below the permissible errors, further analysis of more-DOF systems becomes unnecessary. The 6 types of 2-DOF camera systems as given in Fig. 8 were investigated. For every system, calculations were done to check if the human's head orientations, measured in the experiment of subsection III-C, were representable by the camera system. In case this was not possible, the visualization error was estimated to be the difference between the desired and the closest actual camera orientation. After comparing the visualization errors with the permissible errors, it was found that Type2-1 covers best all the necessary orientations. Since its visualization errors were smaller than the permissible ones we decided to use this system for further analysis.

B. OPTIMIZATION OF TRANSLATIONAL PART

1) 3-DOF camera system: First, camera system (Type2-1) was extended with one translational DOF. Three configurations were considered a DOF in X-, Y- or Z-direction. To make a valid evaluation of the visualization position errors possible, choices for the different link parameters: L_1, L_2 and $\boldsymbol{p}_b$, must be made (see Fig.9). Where $\boldsymbol{p}_b$ is the base position of the 2-DOF camera and L_1 and L_2 relate to the height and the depth from $\boldsymbol{p}_b$ to $\boldsymbol{p}_c$. The location $\boldsymbol{p}_c$ corresponds to the center of the two camera lenses.

Since these parameters partly determine the camera position ($\boldsymbol{p}_c$), their values should be optimized for every possible 3-DOF system. Optimization was done by applying a least square method and minimize the object function V:

$$V = \frac{1}{k}\sum_{i=1}^{k} ||\boldsymbol{p}_h(i) - \boldsymbol{p}_c(i)|| \tag{10}$$

$$\boldsymbol{p}_c(i) = \boldsymbol{p}_b + \begin{pmatrix} \{L_2 \sin\theta_1(i) + L_1 \cos\theta_1(i)\} \cos\theta_2(i) \\ \{L_2 \sin\theta_1(i) + L_1 \cos\theta_1(i)\} \sin\theta_2(i) \\ L_2 \cos\theta_1(i) - L_1 \sin\theta_1(i) \end{pmatrix}, \tag{11}$$

where optimization is done over all the head positions $\boldsymbol{p}_h(i)$, $i = 1 : 100$ that the four subjects realized in the experiment of subsection III-C, θ_1 and θ_2 are the rotations of the 1st (base) and 2nd (upper) rotational joints. V is an average visualization error. After deciding the optimal values for L_1, L_2, $\boldsymbol{p}_b$ for the X-, Y-, Z-translation testbed, the visualization errors of each configuration were calculated and compared to the permissible visualization errors. The left part of Fig.10 shows the visualization errors obtained from the testbed with the the X-translation DOF, showing the best performance of the three configurations. The grey regions in the figure correspond to the permissible error ranges. The resulting performance was found unsatisfactory, consequently we decided to extend the system to 4-DOF.

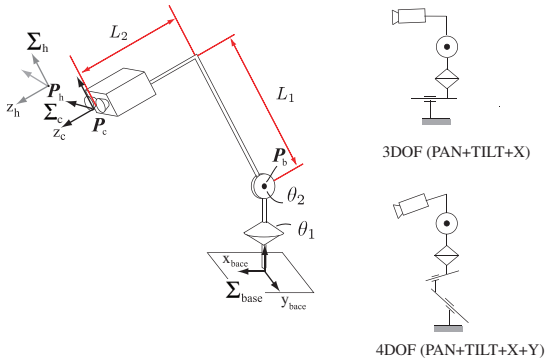


Fig. 9. Model of slave camera system

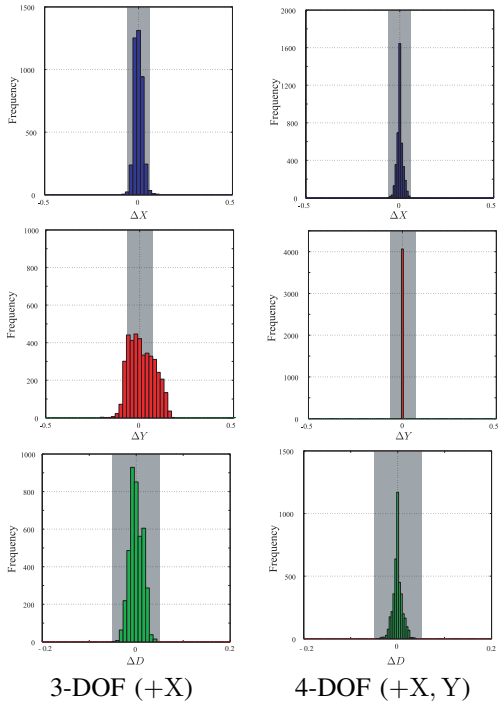


Fig. 10. Histograms for the optimal type for each 3DOF and 4DOF

2) 4-DOF camera system: The rotational part (Type2-1) was extended with two translational DOF's. Among the same lines as previous subsection, a least square optimization of the different parameters was performed for the three possible configurations (XY, XZ, YZ) and visualization errors were computed. It was found that the 4DOF system with XY translation led to an acceptable performance. The right part of Fig.10 shows how the resulting visualization errors remain almost always within the permissible errors. The optimal parameters for this structure were found to be $L_1 = 88.19$mm, $L_2 = 182.81$mm and $\boldsymbol{p}_b = [0.0, 0.0, 512.77]$mm. Based on this calculations, the new camera system was build as depicted on the left part of Fig. 11. The camera is a WAT-231S (Watec Co., CCD) with 13VM2812AS(TAMRON Co.) CCTV lens, used because it is 3 times lighter than the old one and matches the field of view of the HMD completely. The right part of the figure emphasizes the main differences between the rotational part of the original and the new design.

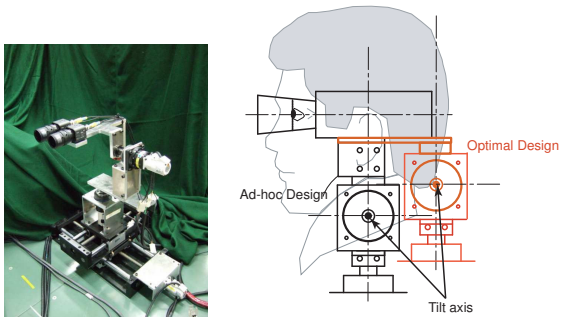


Fig. 11. New design

V. EXPERIMENT

Next the quality of the newly developed camera system was validated. Four subjects were asked to repeat the experiment of II-B, this time using the unified hand/arm master-slave teleoperation testbed and the new camera system. The subjects were allowed to do the experiment only once, so that learning effects do not appear. The quality of the new system is measured by calculating ΔD, ΔX ..., which are measured here as the differences between the operator's head location with respect to the virtual image plane (index m) and the camera system's location with respect to task frame (index s) as $\Delta D = D_m - D_s$ and so on.

Fig.12 shows histograms of the visual errors. The grey backgrounds display the ranges of permissible errors. Appart from a noticable bias that appeared in $\Delta\theta$, ΔY and ΔD (which are caused by some calibration errors,) it can be seen that the range of measured errors exceeds the range of the permissible errors (even after removing the bias). This increase, as compared to simulation, is thought to be caused partly by the servo system not being optimized for the relatively heavy camera system and partly by the ISOTRACK II, which is sampled only at 10Hz.

Ideally speaking also the dynamics of the system should be incorporated in the design. We further need to determine permissible errors for dynamic tasks. In a sense the comparison in Fig.12 is not completely fair as the permissible errors were obtained in a static experiment while the task is a dynamic one. Human's perceptibility in dynamic visualization should be lower than in static tasks. So plotting the 'dynamic permissible errors' on the background of Fig.12 would give a better idea of the quality of the developed system.

VI. CONCLUSION & FURTHER WORK

This paper introduces the concept of permissible visualization errors as a new measure to design higly immersive camera systems for teleoperation. The permissible errors are estimates of the limits of the operator's capability to distinguish visual differences and these errors are task dependent. The new measure not only helps to determine the minimal amount of DOF necessary to create immersive visualization, it also proves to be a good measure to optimize link parameters and configuration.

A redesign of an existing camera system was done based on the new measure. It was found that a system with two rotational (pan, tilt) and two translational axes (x, y) was sufficient to realize the desired tasks with sufficient

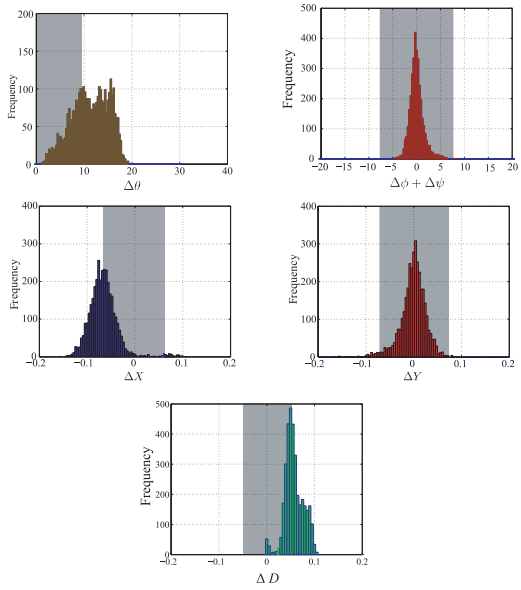


Fig. 12. Histograms for teleoperation experiment

accuracy. In a sense, it is not surprising that the optimal design has two rotational and translational axes. But, we would like to emphasize that this result is reached without any preconceived ideas. Actually, the link lengths in the optimal design can never be found from simply mimicking the appearance of a human neck.

The proposed vision system design framework suggests to break away from simply (and often blindly) mimicking human appearance. Although we only dealt with the vision system design, the same concept could be used for designing any other part of robots.

In the future we will investigate methods to determine 'dynamic permissible errors' to design highly dynamic systems. We will also increase the number of subjects to get a more accurate estimate of the permissible error.

REFERENCES

[1] R.A. Peters and C.L. Campbell, "Robonaut Task Learning through Teleoperation", IEEE International Conference on Robotics and Automation, pp.2806-2811, 2004.

[2] H. Hasunuma et al.: "A Tele-operated Humanoid Robot Drivers a Lift Truck", Proc. of the 2002 IEEE Int. Conf. on Rob. and Autom., pp.2246-2252, 2002.

[3] P. Kammermeier, A. Kron, J. Hoogen and G. Schmidt: "Display of Holistic Haptic Sensations by Combined Tactile and Kinesthetic Feedback", Presence, Vol.13, No.1, pp.1-15, 2004.

[4] C. Preusche et al.: "Flexible Multimodal Telepresent Assembly using a Generic Interconnection Framework", Proc. of the 2002 IEEE Int. Conf. on Rob. and Autom. pp.3712-3718, 2002.

[5] S. Tachi and K. Yasuda: "Evaluation Experiments of a Teleexistence Manipulation System", Presence, Vol.3, No.1, pp.35-44, 1994.

[6] B.G. Brooks and G.T. McKee: "The Visual Acts Model for Automated Camera Placement During Teleoperation", IEEE International Conference on Robotics and Automation, pp.1919-1924, 2001.

[7] B. Dejong, E. Colgate, E. Peshkin: "Improving Teleoperation: Reducing Mental Rotations and Translations",IEEE International Conference on Robotics and Automation, pp.3708- 3714, 2004.

[8] H. Baier, M. Buss, F. Freyberger, and G. Schmit: "Interactive stereo vision telepresence for correct communication of spatial geometry", Advanced Robotics, Vol.17, No.3, pp.219-233, 2003.

[9] Y. Yokokohji, Y. Iida and T.Yoshikawa: "'Toy Problem' as the Benefits of Combined Active Stereo Vision and Haptic Telepresence", Proc. of the 2000 IEEE/RSJ Int. Conf. on Intell. Rob. and Autom., pp.3173-3178, 2000.

[10] R. P. Paul: Robot Manipulators, MIT Press, Cambridge, Mass., 1981