1. Context

1.1 Research framework : Embodied Learning

Our goal is to “teach” a robot to interact autonomously in a face-to-face task with a human. Due to the complexity of the task, standard learning approach like learning by observation and kinesthetic demonstration are not efficient enough notably for social signals (such as gaze or head movements). Our approach, based on the embodied learning paradigm, will teach the robot with his own moves by embodied him [10]. Like a puppeteer, a pilot controls the robot remotely using an immersive teleoperation platform. In order to record an interaction with minimal bias, the platform should become “transparent” and the remote user will teach the robot with his own moves by embodied him [10]. Like a puppeteer, a pilot will teach the robot with his own moves by embodied him [10].

1.2 State of the Art

What we know about humans :

- Depth perception is based on several factors : Binocular disparity (stereovision), occlusion, parallax, convergence, known semantics of the objects, [16, 5].
- Stereoscopic vision is useful before 15m (cannot differentiate from monovision after)[1].
- Vvergence is useful in the peri-personal space (2+0m)[14].

Current use of immersive teleoperation

- Search and rescue robot [8]
- Drone navigation [4]
- Immersive telepresence [6, 7, 3]

Gaze controlled methods

Virtual gaz joystick : “Moving to the center” [17][12]
EyeSeeCam [11, 13]

Limitations of current immersive teleoperation devices

- Underestimation of depth in peripersonal space (2m), overestimation after 2m [2].
- Sensory motor conflict: Accommodation-convergence conflict [15].

Drawback : In those setups, the stereo rig is fixed. The pilot loses the vergence information control, has reduced depth perception and experiences the accommodation-convergence conflict. For a human facing the robot, the robot gaze is less interpretable.

1.3 SGCS : Stereo Gaze Contingent Steering

Proposed approach : Here we propose a new natural control method for a pair of stereoscopic robotic eyes with vergence abilities, called SGCS (Stereo Gaze Contingent Steering), running alongside the control of a robotic head. An evaluation of the control method has also been performed.

2. Technological platform

We use the Mical platform “NINA” from the CRISSP team at Gipsa. Specification :

- Icos 2.0 with enhanced face articulation [9].
- Cluster of 4 PC (3 Linux + 1 Windows) running a yarp (client-server robotic middleware).
- HTC Vive + SMI integration for eye-tracking
- IPD cameras equivalent to human IPD => reduced hyperstereopsis
- Communication with UDP/TCP through the YARP middleware
- The camera’s feeds are synchronised and displayed in the HMD (Head Mounted display) as video texture.

3. Control methods

The control of the head and eye is done trough a angular command for the six head encoders:

\[ \text{U Gaze} = \text{U neck} \times \text{U roll} \times \text{U pitch} \times \text{U yaw} \text{ eyes} \times \text{U eyes tilt} \times \text{U eyes depth} \]

1. Head control : The head angles are driven by the HMD orientation value (standard approach).

2. Eye control :

(a) The gaze information returned (in pixel) is converted in UV coordinates, relative to the displayed video texture referential.
(b) StereoClick : Using a transfer matrix \( UV_{to\_angle} \) (inverse model), the model is able for a chosen target (defined by a UV pair) to return an absolute angular command value to center the two stereo cameras images on it. This inverse linear model has (surprisingly) a precision of 0.3 on the three angles.

\[ UV_{to\_angle} = \left[ \begin{array}{ccc} u & v & 1 \end{array} \right] \begin{bmatrix} L_{x} & R_{y} & R_{z} \end{bmatrix} = [\text{left\_eyes} \text{right\_eyes}] \]

3. Foveal display : Move the center of the video texture in the virtual head to a new coordinates pair, calculated by the forward model with the eyes encoders angular values. This moves the video texture in the virtual world to a coherent position for the robot and cues on the pilot side.

4. Platform validation

4.1 Setup & protocol

Setup : 7 target at various distance (25 to 100cm)
Protocol : For the reference command (ideal target angles determined semi-automatically) and the pilot, every target has been seen 8 times. On 4 passes (left->right, front->back, right->left, back->front) repeated two times.
Subjects : 13 subjects (3 women, 13 men), aged between 22-36 y.o. No prior experience of virtual reality before for most of them (13 unexperienced VR).

4.2 Results

Discussion : Our SGCS control method is able to move the robotic eye in coherence with the orientation of the human eye (the camera optical axis is aligned with the human gaze). The cameras are looking where the human is looking with respect of tilt, azimut, and vergence.

Future works :
- Hypothesis : Control of vergence improves perception and evaluation of depth in the near and medium field while maintaining oculomotor cues and reducing the accommodation-convergence conflict.
- Improve the reactivity of the control method : detection of fixation and saccade.

5. Discussion & future works

References

[5] James R. Tresilian, Mark Mon-Williams, and Benjamin M. Kelly. Increasing confidence in spatial orientation of the human eye (the camera optical axis is aligned with the human gaze). The cameras are looking where the human is looking with respect of tilt, azimut, and vergence.

Future works : Hypothesis : Control of vergence improves perception and evaluation of depth in the near and medium field while maintaining oculomotor cues and reducing the accommodation-convergence conflict.

- Improve the reactivity of the control method : detection of fixation and saccade.