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 approaches like learning by observation and kinesthetic demonstration are not efficient, notably for demonstrating social signals (such as gaze or head movements). Our approach consists in exploring the embodied learning paradigm, where a human pilot teaches the robot with his/her own moves [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 world represented in a “natural” way. Our research aims at providing the pilot with an active perception of the remote space, notably with a trustful and coherent perception of depth.

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]
- Vergence 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 gaze 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].

Drawback: In those setups, the stereo rig is fixed. The pilot loses the vergence information and kinesthetic demonstration are not efficient, notably for demonstrating social signals (such as gaze or head movements). Our approach consists in exploring the embodied learning paradigm, where a human pilot teaches the robot with his/her own moves [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 world represented in a “natural” way. Our research aims at providing the pilot with an active perception of the remote space, notably with a trustful and coherent perception of depth.

3. Control methods

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

\[
\begin{align*}
\text{angle}_{\text{neckL}} & = \text{pitch}_L - \text{tilt}_L + \pi/2 \\
\text{angle}_{\text{neckR}} & = \text{pitch}_R - \text{tilt}_R - \pi/2
\end{align*}
\]

1. Head control: The head angles are driven by the HMD orientation value (standard approach).
2. Eye control: The gaze information returned (in pixel) is converted in UV coordinates, relative to the displayed video texture referential.
3. Inverse model: Using a transfer matrix UV_to_angle ~ that links the ROI placement on the rectified stereo camera images with the angular values that would align the ROI with the cameras’ optical axis –, this inverse model carries over the pilot’s gaze direction to the robot’s eyes movements. This inverse linear model has (surprisingly) a precision of 0.5° on the three angles.

4. Foveal display: We then move the center of the video texture to a new UV coordinates pair, calculated by the Forward model from 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.

4.1 Setup & protocol

Setup: 7 target at various distance (25 to 100cm)

Protocol: For the reference condition (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: 16 subjects (3 women, 13 men), aged between 22-56 yo. No prior experience of virtual reality before for most of them (13 inexperienced 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 cameras’ optical axes are aligned with the human gaze).

The cameras are looking where the human is looking with respect of tilt, azimuth, 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-vergence conflict.
* Improve the reactivity of the control method: detection of fixation and saccade.

References

[1] Remi CAMBUZAT, Frederic ELISEI, Gerard BAILLY

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab/DPC, Grenoble, France

{remi.cambuzat,frederic.elisei,gerard.bailly}@gipsa-lab.fr

SGCS: Stereo Gaze Contingent Steering for Immersive Telepresence

Grenoble Images Parole Signal Automatique

UMR CNRS 5225 - Grenoble Communication et Image, CNRS, Grenoble INP, GIPSA-lab/DPC, Grenoble, France

38400 Saint Martin d’Hères - FRANCE

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, notably for demonstrating social signals (such as gaze or head movements). Our approach consists in exploring the embodied learning paradigm, where a human pilot teaches the robot with his/her own moves [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 world represented in a “natural” way. Our research aims at providing the pilot with an active perception of the remote space, notably with a trustful and coherent perception of depth.

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]
- Vergence 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 gaze 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].

Drawback: In those setups, the stereo rig is fixed. The pilot loses the vergence information and kinesthetic demonstration are not efficient, notably for demonstrating social signals (such as gaze or head movements). Our approach consists in exploring the embodied learning paradigm, where a human pilot teaches the robot with his/her own moves [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 world represented in a “natural” way. Our research aims at providing the pilot with an active perception of the remote space, notably with a trustful and coherent perception of depth.

3. Control methods

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

\[
\begin{align*}
\text{angle}_{\text{neckL}} & = \text{pitch}_L - \text{tilt}_L + \pi/2 \\
\text{angle}_{\text{neckR}} & = \text{pitch}_R - \text{tilt}_R - \pi/2
\end{align*}
\]

1. Head control: The head angles are driven by the HMD orientation value (standard approach).
2. Eye control: The gaze information returned (in pixel) is converted in UV coordinates, relative to the displayed video texture referential.
3. Inverse model: Using a transfer matrix UV_to_angle ~ that links the ROI placement on the rectified stereo camera images with the angular values that would align the ROI with the cameras’ optical axis –, this inverse model carries over the pilot’s gaze direction to the robot’s eyes movements. This inverse linear model has (surprisingly) a precision of 0.5° on the three angles.

4. Foveal display: We then move the center of the video texture to a new UV coordinates pair, calculated by the Forward model from 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.

4.1 Setup & protocol

Setup: 7 target at various distance (25 to 100cm)

Protocol: For the reference condition (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: 16 subjects (3 women, 13 men), aged between 22-56 yo. No prior experience of virtual reality before for most of them (13 inexperienced 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 cameras’ optical axes are aligned with the human gaze).

The cameras are looking where the human is looking with respect of tilt, azimuth, 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-vergence conflict.
* Improve the reactivity of the control method: detection of fixation and saccade.

References