SGCS: Stereo Gaze Contingent Steering for Immersive Telepresence
Rémi Cambuzat, Frédéric Elisei, Gérard Bailly

To cite this version:

HAL Id: hal-01638383
https://hal.archives-ouvertes.fr/hal-01638383
Submitted on 21 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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 world represented in a "natural" way. Our actual research aims at rendering a coherent representation of the remote space and depth perception.

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/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, SGCS (Stereo Gaze Contingent Steering), running alongside the control of a robotic head. An evaluation of the control method has already been performed.

2. Technological platform

We use the Mical platform Nina from the CRISSP team at Gipsa. Specification:

- iLocs 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
- Head-mounted display for immersive 3D teleoperation of field robots.

Platform process

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

$$\mathbf{UV}_{\text{target}} = \mathbf{UV}_{\text{current}} + \Delta\mathbf{UV}_{\text{command}}$$

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 reference.
   (b) StereoClock: Using a transfer matrix $$\mathbf{UV}_{\text{to_angle}}$$ (inverse model), the module 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 0.5 on the three angles.

3. Foveal display: Move the center of the video texture in the virtual environment to a new UV 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.