SGCS: Stereo Gaze Contingent Steering for Immersive Telepresence

Rémi CAMBUZAT, Frédéric ELISEI, Gérard BAILLY
Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab/DPC, Grenoble, France
{remi.cambuzat,frederic.elisei,gerard.bailly}@gipsa-lab.fr

JJCR'17 Bidart 64

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]

Our beaming platform "NINA" [9, 10]

Gaze controlled methods

- Virtual gaze joystick : "Moving to the center" [17] [12]
- EyeSeeCam [11, 13]

Limitations of current immersive teleoperation devices

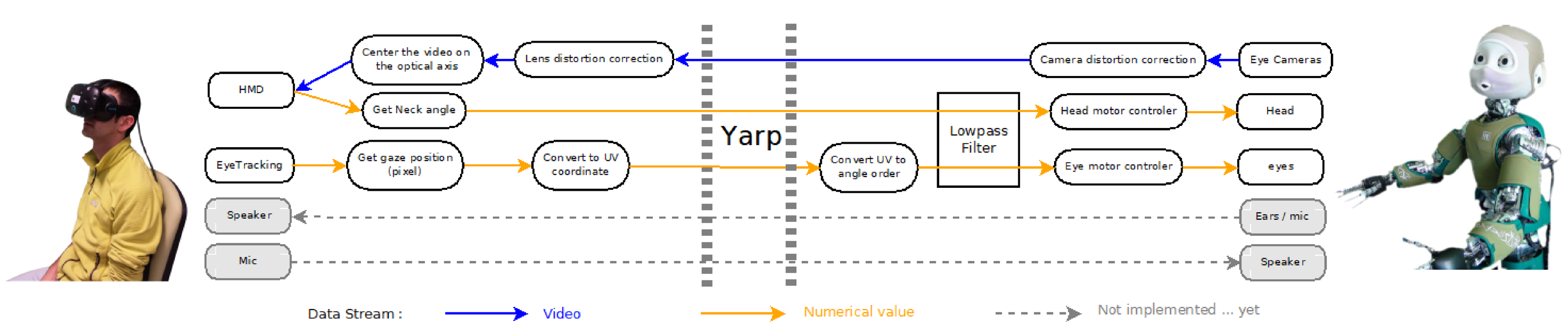
- Underestimation of depth in peripersonal space (<2m), overestimation after 2m [2].
- Sensory motor conflict: Accommodation-convergence conflict [15]

Drawback : In those setups, the stereo rig is fixed. The pilot loses the vergence information/control, has reduced depth perception and experiences the accomodation-vergence 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



Beamning process

We use the Mical platform *NINA* from the CRISSP team at Gipsa. Specification :
* Icub 2.0 with enhanced face articulation [9].
* Cluster of 4 PC (3 Linux + 1 windows) using 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

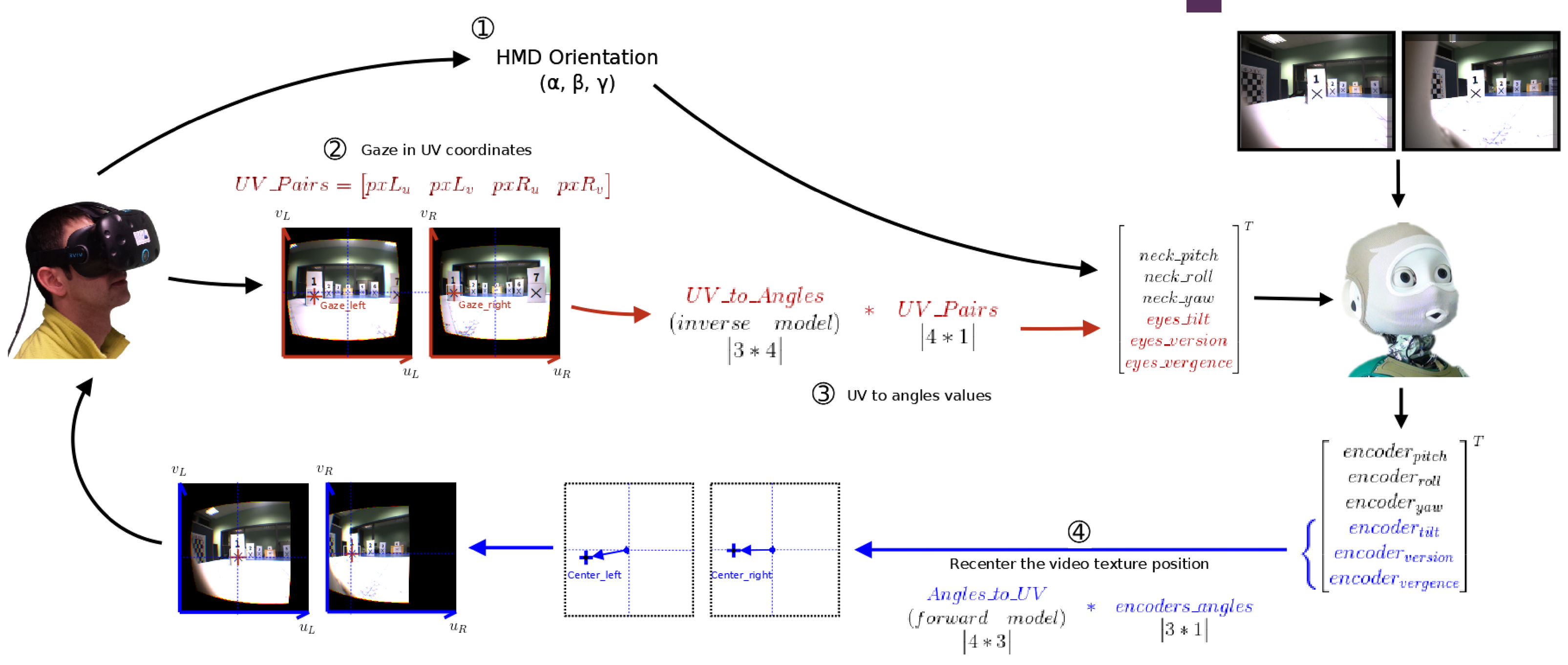
The control of the head and eye is done trough a angular command for the six head encoders:

$$[neck_{pitch}\ neck_{roll}\ neck_{yaw}\ eyes_{tilt}\ eyes_{version}\ eyes_{vergence}]$$

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.

$$UV\_to\_angle * [pxL_u\ pxL_v\ pxR_u\ pxR_v] = [eyes_{tilt}\ eyes_{version}\ eyes_{vergence}]$$

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.



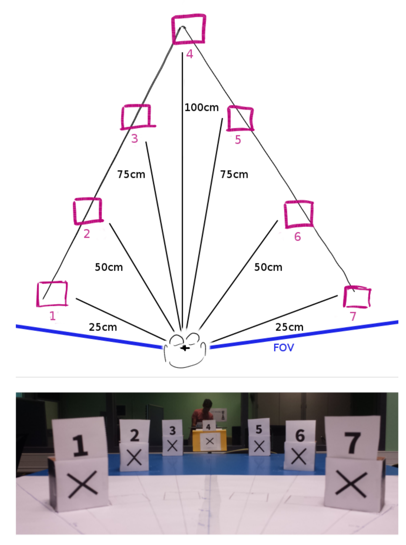
Overall controls methods

4. Platform validation

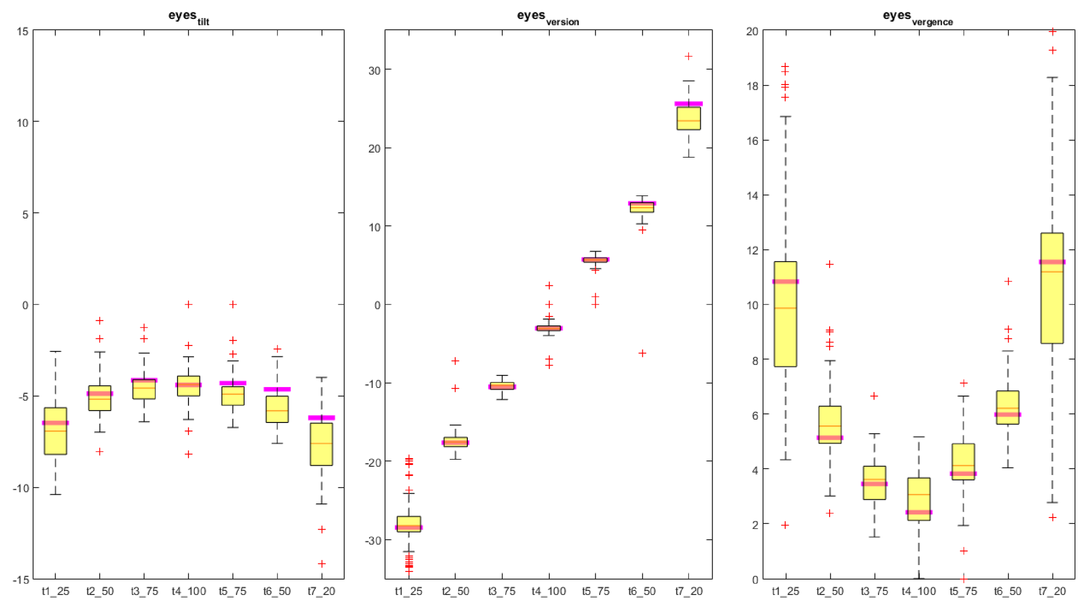
4.1 Setup & protocol

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).

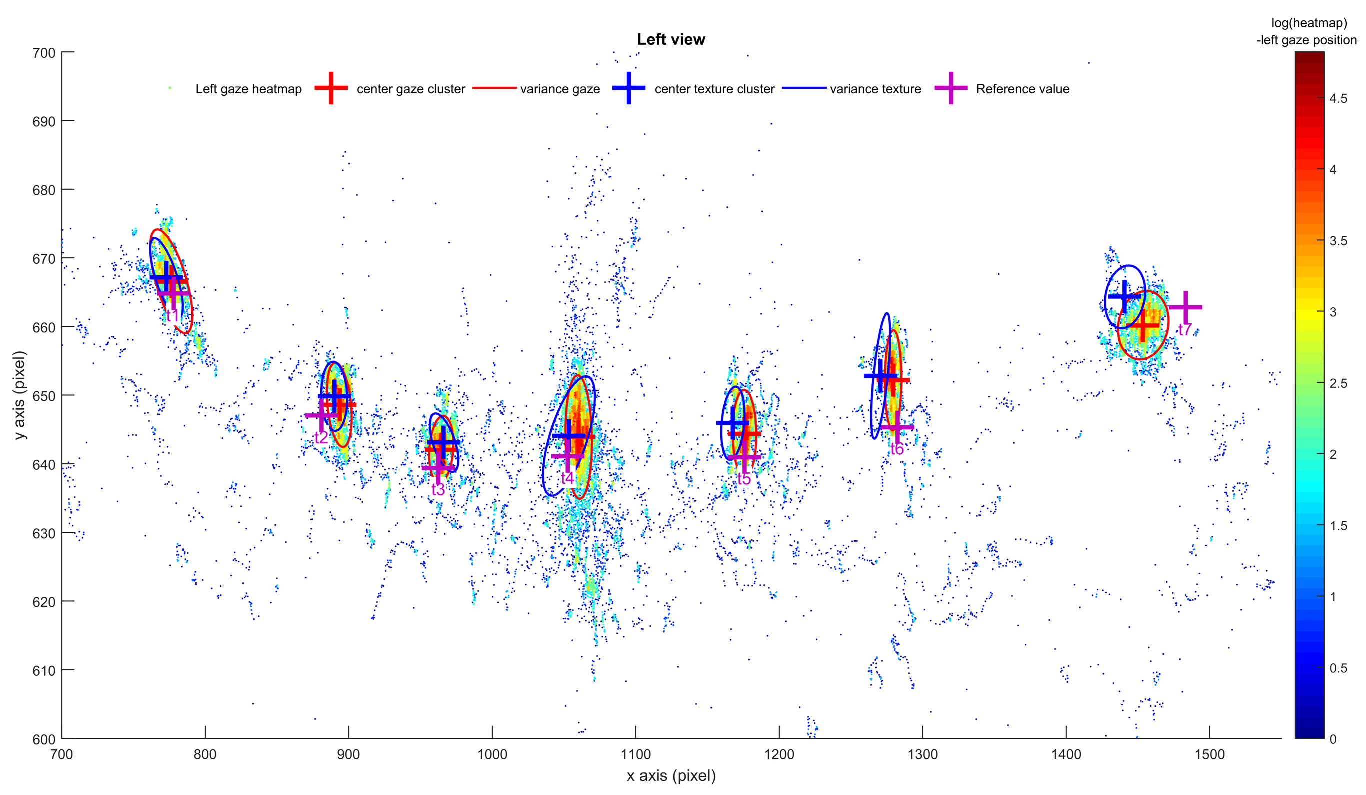
4.2 Results



SGCS characterisation setup



Encoders angle values for all subjects.The reference is in magenta



Heat map generated from the left eye position and associated visual feedback positions in pixels (one subject)

The *motion-to-photon* mesured latency is ~200ms.

5. Discussion & future works

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 accomodation-vergence conflict.
* Improve the reactivity of the control method: detection of fixation and saccade.

References

[1] Robert S Allison and Barbara J Gillam. Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, 91010(1110101), 2009.

[2] C Armbrüster, M Wolter, T Kuhlen, W Spijkers, and B Fimm. Depth perception in virtual reality: distance estimations in peri- and extrapersonal space. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, 11(1):9–15, 2008.

[3] Doraplatform. DORA Platform, 2016.

[4] Antonio Fernández-Caballero, Yudong Zhang, Nikolai Smolyanskiy, and Mar Gonzalez-Franco. stereoscopic First Person View system for Drone navigation. 4(11), 2017.

[5] Kelly S. Hale and Kay M. Stanney. Effects of low stereo acuity on performance, presence and sickness within a virtual environment. *Applied Ergonomics*, 37(3):329–339, may 2006.

[6] Sven Kratz and Fred Rabelo Ferriera. Immersed Remotely: Evaluating the Use of Head Mounted Devices for Remote Collaboration in Robotic Telepresence.

[7] Uriel Martinez-Hernandez, Michael Szollosy, Luke W Boorman, Hamideh Kerdegari, and Tony J Prescott. Towards a wearable interface for immersive telepresence in robotics.

[8] Henrique Martins, Ian Oakley, and Rodrigo Ventura. Design and evaluation of a head-mounted display for immersive 3D teleoperation of field robots. *Robotica*, 33(10):2166–2185, dec 2015.

[9] Alberto Parmiggiani, Marco Randazzo, Marco Maggiali, Frederic Elisei, Gerard Bailly, and Giorgio Metta. An articulated talking face for the iCub. In *IEEE-RAS International Conference on Humanoid Robots*, volume 2015-Febru, pages 1–6, 2015.

[10] Miquel Sauze and Gérard Bailly. Beaming the Gaze of a Humanoid Robot. 2015.

[11] Erich Schneider, Thomas Villgrattner, Johannes Vockeroth, Klaus Bartl, Stefan Kohlbecher, Stanislavs Bardins, Heinz Ulbrich, and Thomas Brandt. EyeSeeCam: An Eye Movement-Driven Head Camera for the Examination of Natural Visual Exploration. *Annals of the New York Academy of Sciences*, 1164(1):461–467, may 2009.

[12] Sophie Stellmach and Raimund Dachselt. Designing Gaze-based User Interfaces for Steering in Virtual Environments fland Raimund Dachselt y Sophie Stellmach User Interface & Software Engineering Group Faculty of Computer Science University of Magdeburg , Germany. pages 131–138, 2012.

[13] Josef Stoll, Stefan Kohlbecher, Svenja Marx, Eric Schneider, and Wolfgang Einhäuser. Mobile three dimensional gaze tracking (PDF Download Available).

[14] James R Tresilian, Mark Mon-Williams, and Benjamin M Kelly. Increasing confidence in vergence as a cue to distance.

[15] Margarita Vinnikov and Robert S. Allison. Gaze-Contingent Depth of Field in Realistic Scenes: The User Experience, 2014.

[16] J.E. Cutting Vishton and P.M. chapter Perceiving Layout and Knowing Distances : The Integration, Relative Potency, and Contextual Use of Different Information about Depth. *Perception of Space and Motion*, 22(5):69–117, 1995.

[17] Dingyun Zhu, Tom Gedeon, and Ken Taylor. "Moving to the centre": A gaze-driven remote camera control for teleoperation. *Interacting with Computers*, 23(1):85–95, jan 2011.

Grenoble Images Parole Signal Automatique
UMR CNRS 5216 - Grenoble Campus
38400 Saint Martin d'Hères - FRANCE