

Inter-facial relations: Binocular geometry when eyes meet

Nicholas Wilkinson

Robotics, Brain, and Cognitive Sciences
Istituto Italiano di Tecnologia
Via Morego, 30, Genova 16122
nicholas.wilkinson@iit.it

Giorgio Metta

Robotics, Brain, and Cognitive Sciences
Istituto Italiano di Tecnologia
Via Morego, 30, Genova 16122

Gustaf Gredebäck

BabyLab
Department of Psychology
Uppsala University
Sweden

Abstract—Infants are sensitive to eye contact and face like stimuli from birth. We advance the novel hypothesis that the observed sensitivity of human neonates to eye contact may be mediated in part by sensor distribution. The arrangement of the observers eyes acts as a morphological template for eye-like feature pairs in the world, providing certain geometric constraints on stereo relations which may help to focus attention on eye-like stimuli. We give a brief proof-of-concept demonstration and discuss how this morphological mechanism could be exploited by humans and humanoid robots to facilitate the establishment and maintenance of eye contact. We advance three experimental hypotheses which would help to distinguish this mechanism from existing explanations.

Keywords—embodiment, sensor distribution, binocular vision, face detection

I. INTRODUCTION

Human infants are sensitive to eye contact [4] and face-like stimuli [8] from birth. Currently favoured explanations are based on generic features of the visual system [6] in combination with a crude sub-cortical face template - the “CONSPEC” mechanism [12]. Here we suggest an additional, non-exclusive mechanism based on sensor distribution. We are exploring the role of sensor distribution in ongoing “perceptual crossing” [1] studies on the iCub robot (icub.org).

II. INTER-FACIAL RELATIONS

We make the following assumptions to simplify the explanatory process. These assumptions place powerful constraints on geometric relations between pairs of eyes. Although they do not hold in the real world, the mechanism is robust to deviation from the idealised case at the expense of increased computation and/or decreased reliability.

A. Simplifying assumptions

- 1: All human faces have the same inter-eye distance.
- 2: The gaze directions of left and right eyes are parallel.

B. The rectangle of mutual gaze

We consider first the “mutual gaze” case where the eyes are directly looking at each other (Figure 1a). Our assumptions imply that the points of the four eyes lie at the corner positions of a rectangle. The retinal locations of opposite eyes overlap at the retinal centre $R1=L2$ in the stereo pairs. A mirror symmetry also maps $L1$ to $R2$. The existence of the overlap

at the retinal centre plus the symmetry will thus be a useful indicator of meeting another pair of eyes in mutual gaze. These sensory events occur simply because the size and form of the subject’s sensory morphology matches that of the object’s.

C. Detecting the binocular overlap

The human eye, with its dark centre and white sclera, is particularly suited to centre-surround contrast detection [9] and the iCub robot shares this highly detectable facial feature. Centre surround (henceforth “CS”) connectivity is widely observable in the visual system. Simple salience calculation based on multi-scale centre-surround filters will tend to assign high monocular salience to eye stimuli. Thanks to the $R1=L2$ overlap, monocular salience can be amplified by combining the salience maps for each eye through summation or multiplication. See Figure 2 for an example. This may be sufficient to bias neonate orientation movements towards high CS contrast features at eye distance apart. As binocular visual circuitry develops, the overlap could be detected and tracked more precisely as a localised false match in disparity calculations. Poor vergence movement will disrupt the overlap mechanism. Good vergence recovers a different set of overlap relations.

D. Deviations from the perfectly overlapping case

Individual differences, postural variation and the noise of the real world mean that perfect zero-disparity overlaps are unlikely in practice. Individual differences in inter-eye distance will introduce spatial deviations which will usually be small and linearly proportional. Rotation could introduce large deviations. However, where head rotation is preceded by mutual gaze, the overlap could be identified and then tracked to infer the rotation. This might help to explain the importance of initial eye contact in establishing social interactions such as joint attention [5]. In general, deviations from the idealised perfect overlap case will incur computational costs and/or reduced reliability proportional to the extent of the deviation.

E. The parallelogram of translation and the trapezoid of rotation

The overlap $R1=L2$, though not its retinal location or size, is invariant to translations from the mutual gaze case given our assumptions. Inter-eye distance is assumed to be identical, and

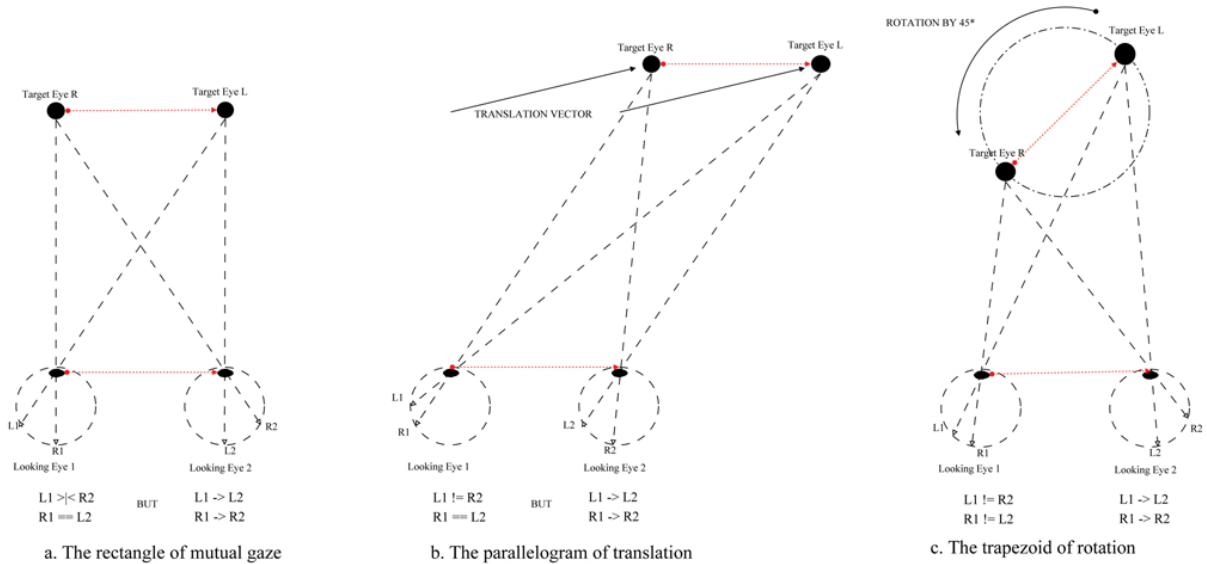


Fig. 1. (a) The symmetric case - mutual gaze. There is one zero disparity overlap, and a mirror symmetry. (b) The zero disparity overlap remains under translation. (c) The zero disparity overlap is destroyed by rotation, as apparent inter-eye distance changes.

so the four eyes will lie at the corners of a parallelogram under any translation (Figure 1b). This preserves the parallel relation between $R1$ and $L2$ and hence also preserves the overlap. Translations in depth affect the retinal size of the overlapping features, while lateral translations affect the retinal location of the overlap.

Rotation of the head (1c) will cause deviations from perfect overlap which will depend on the amplitude of rotation. This is consistent with, and could partially explain, findings that detecting mutual gaze is more difficult under rotations. Of course a boundary exists where one eye goes out of view.

III. DEMONSTRATION

Figure 2 depicts a simple proof-of-concept demonstration of the mechanism. A monocular image of the iCub robot (2c, underlaid) is split into a pseudo-stereo pair. This is reasonable because left and right gaze are assumed to be parallel and the image has little relevant occlusion. We assume an inter-ocular distance of 130 pixels. This was chosen by hand based on the approximate distance separating the eyes of the iCub in the image. Though this may seem like a cheat, it is in fact precisely the role played by sensor distribution. A multi-scale CS filter is applied to the stereo images, producing two contrast based salience maps (2d,e). Thanks to the $R1$ to $L2$ approximate overlap, the pointwise product of 2d and 2e amplifies the salience of the overlapping eyes, whilst reducing that of other regions (2f). Note that *any* CS-salient features at eye distance apart will be amplified in this way.

IV. DISCUSSION

Stereo vision and embodiment constraints have been applied to face detection elsewhere [10],[2]. Disparity mismatch has been used to model other visual phenomena [7]. We are not aware of existing explanation of the specific mechanism

described here, though it is difficult to imagine that the proposal is entirely novel. It is worth noting that in 2D images depth is constant and inter-eye distance varies, but in 3D inter-eye distance is constant and depth varies, so the geometric relations are very different.

No pixelwise iteration of pattern kernels across the whole image, and no stimulus specific experience or information storage is required. The morphology of the lookers' eyes embodies a rough description of the target eyes. Sensor distribution is enough to increase the focus of standard visual processing on eye-like stimuli. The computational trade off is the level of deviation from the zero disparity case one wishes to accommodate. Such deviations will be information rich, reflecting individual differences, translation and rotation according to well defined geometric principles. Direct eye contact provides the most powerful and easy to detect cue - near perfect overlap plus mirror symmetry. From here, changes can be tracked as they occur, providing useful information about mutual orientation with the interactor. This could help to explain the importance of initial eye contact in establishing a social interaction [5].

It is known that inversion greatly reduces sensitivity to faces [3]. This cannot be explained by visual sensor distribution alone. However, we can suggest an extension to address this. Prenatal haptic exploration of the face establishes the position of the mouth relative to the eyes. Visual detection of an eye-like feature pair triggers associative mechanisms which suggest the point in the image, relative to the eye-like features, where the mouth should be. The visual input from this part of the scene is copied to sensory and/or motor surfaces associated with the mouth, and tested for fit with previous experience. A good match suggests a mouth, and hence a face. This embodied, multi-modal style of perception may also help to explain neonatal imitation of facial gestures [11], a possibility

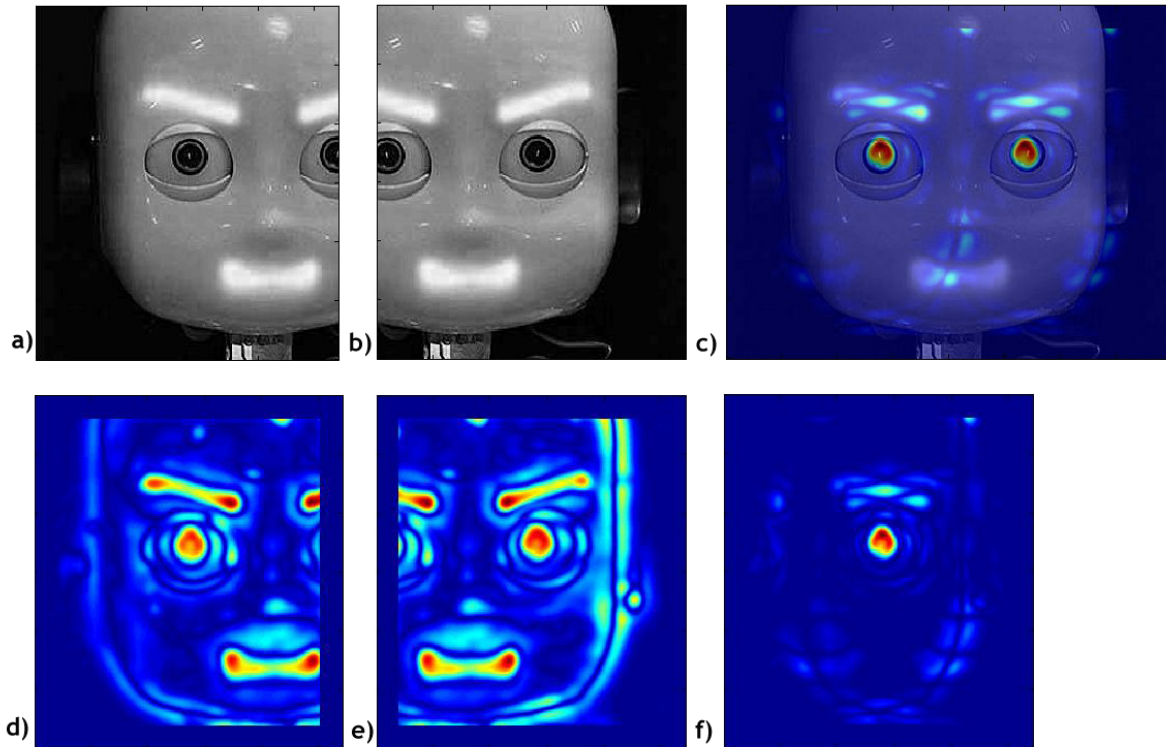


Fig. 2. A demonstration of how sensor distribution can focus attention on eyes. (a) and (b) are the left and right pseudo-stereo pair derived from a single image of the iCub robot. (d) and (e) show the summed output of multi-scale centre-surround filtering. (f) depicts the pointwise product of (d) and (e). (c) shows (f) projected on to the original monocular image wrt left and right eye positions.

we will investigate on the iCub robot.

We can propose three experimental hypotheses which would help to disentangle the role of binocular sensor distribution relative to existing (monocular) theories of neonate face preference.

H1 - Neonate face preference will decrease when the face stimulus used is larger and farther away, preserving monocular retinal image features but not binocular relations. So any behavioural change can be attributed to binocular mechanisms.

H2 - The disappearance and reappearance of neonatal face preference [8] will correlate with learning of vergence movement in individual infants.

H3 - Similar inter-eye distance will facilitate and improve social interaction experience between individuals of any age.

ACKNOWLEDGMENT

This research is being supported by the EU project Robot-Doc under 25065 from the 7th Framework Programme, Marie Curie Action ITN.

REFERENCES

- [1] M. Auvray, C. Lenay, and J. Stewart. Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, 27(1):32 – 47, 2009.
- [2] M. Dixon, F. Heckel, B. Smart, and R. Pless. Faster and more accurate face detection on mobile robots using geometric constraints. *Proc. IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2007.
- [3] T. Farroni, G. Csibra, and M. Johnson. Mechanisms of eye gaze perception during infancy. *Journal of Cognitive Neuroscience*, 16:1320–1326, 2004.
- [4] T. Farroni, G. Csibra, F. Simion, and M. Johnson. Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, 99:9602–9605, August 2002.
- [5] T. Farroni, D. Pividori, F. Simion, S. Massaccesi, and M. Johnson. Eye gaze cueing of attention in newborns. *Infancy*, 5(1):39–60, 2004.
- [6] T. Farroni, E. Valenza, F. Simion, and C. Umiltà. Configural processing at birth: evidence for perceptual organisation. *Perception*, 29:355–372, 2000.
- [7] G. Francis. Cortical dynamics of figure-ground segmentation: Shine-through. *Vision Research*, 49(1):140 – 163, 2009.
- [8] M. Johnson, S. Dziurawiec, H. Ellis, and J. Morton. Newborns preferential tracking of faces and its subsequent decline. *Cognition*, 40(1-2):1–19, August 1991.
- [9] H. Kobayashi and S. Kohshima. Unique morphology of the human eye and its adaptive meaning: comparative studies on the external morphology of the primate eye. *Journal of Human Evolution*, 40(5):419–435, May 2001.
- [10] S. Kosov, K. Faber, T. Thormalen, and H. Seidel. Rapid stereo enhanced face detection. *Proceedings 16th IEEE International Conference on Image Processing*, pages 1221–1224, November 2009.
- [11] A. Meltzoff and K. Moore. Newborn infants imitate adult facial gestures. *Child Development*, 54(3):702–709, June 1983.
- [12] J. Morton and M. Johnson. Conspic and conlern: A two-process theory of infant face recognition. *Psychological Review*, 98(2):164–181, April 1991.