Recovering depth from stereo without using any oculomotor information

Tadamasa Sawada
School of Psychology, National Research University Higher School of Economics

The human visual system uses binocular disparity to perceive depth within 3D scenes. It is commonly assumed that the visual system needs oculomotor information about the relative orientation of the two eyes to perceive depth on the basis of binocular disparity. The necessary oculomotor information can be obtained from an efference copy of the oculomotor signals, or from a 2D distribution of the vertical disparity, specifically, from the vertical component of binocular disparity. It is known that oculomotor information from the efference copy and from the vertical disparity distribution can affect the perception of depth based on binocular disparity. But, these effects are too slow and too unreliable to explain the stable and reliable depth perception we have under natural viewing conditions when natural eye movements are made. This study describes a computational model that recovers depth from a stereo-pair of retinal images without being given any oculomotor information.

The model recovers the depth within a 3D scene from a stereo-pair of its retinal images. Each retinal image is represented by a set of visual angles between pairs of points in the scene. The depth recovered is represented in a head-centric coordinate system, except when a rotation is made around the interocular axis between the two eyes. The representations of the retinal images and the recovered depth, as well as the process used to recover depth, do not vary with eye movements.

The model recovers the depth of the scene by using a 2D optimization method. Each 3D scene is recovered so as to make the stereo-pairs of lines-of-sight intersect as well as possible within the recovered scene. The space of this optimization is characterized by the shape of a triangle formed by the arbitrary selection of any 3 points within the scene.

The mathematical validity and the computational robustness of the model were tested in a simulation experiment. Results of this simulation show that the model can recover depth within a 3D scene from a stereo-pair of its 2D retinal images veridically and reliably when there are 6, or more than 6 points, in the scene.

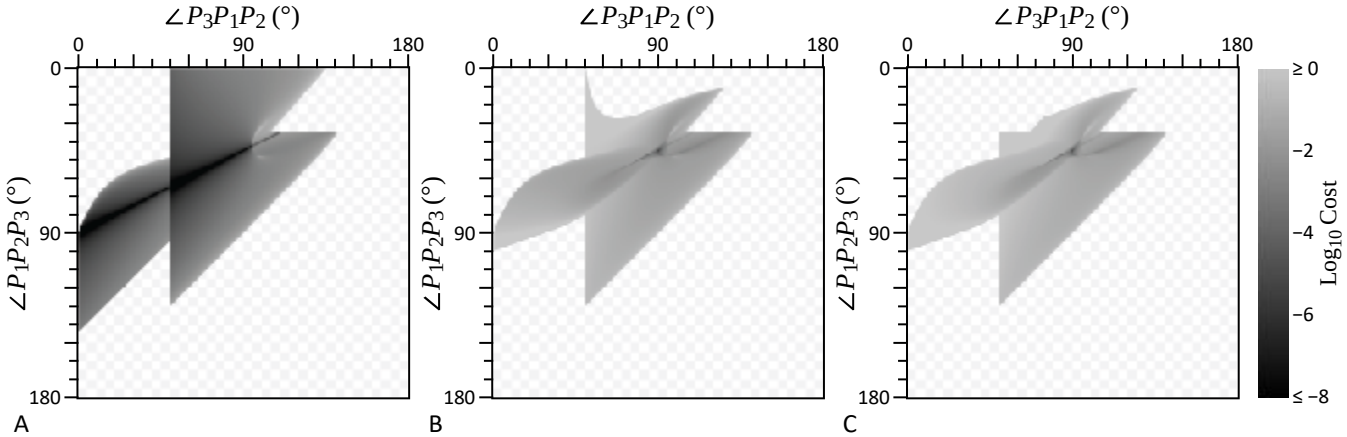

Figure. 2D distributions of the cost computed from stereo-pairs of the retinal images of 3D scenes (A) with 4 points ($P_1 = [-20\ -20\ 60]^t$, $P_2 = [-20\ 20\ 60]^t$, $P_3 = [20\ -20\ 60]^t$, $P_4 = [20\ 20\ 60]^t$), (B) with the 5th point $P_5 = [-10\ 0\ 40]^t$, and (C) with the 6th point $P_6 = [10\ 0\ 40]^t$. This scene was viewed from a stereo-pair of eyes at $[-3.3\ 0\ 0]^t$ and $[3.3\ 0\ 0]^t$. The abscissa and ordinate of these graphs represent $\angle P_3P_1P_2$ and $\angle P_1P_2P_3$ of the triangle T_{123} formed by P_1, P_2, and P_3. The cost is indicated by grayscale levels. The smaller the cost value, the better the stereo-pairs of lines-of-sight intersect with one another within a 3D scene. The checkered regions indicate that a valid 3D scene cannot be recovered with the shape of T_{123} in these regions. The global minimum of the distribution can be specified uniquely at ($\angle P_3P_1P_2$, $\angle P_1P_2P_3$) = (90°, 45°) if there are 5 or 6 points within the scene.