Representations of scene statistics in the primary visual cortex for inferring binocular disparity.

Determining depth and identifying 3D shape from images is a difficult problem that our visual system handles very efficiently. Features, structures, and patterns in an image can have numerous potential 3D interpretations, which necessitates that depth perception is solved by inference using a multitude of visual cues to gather as much evidence as possible. The relative depth of any feature with respect to fixation can be determined by triangulating the horizontal shift or disparity between the images of this feature projected onto the left and the right eyes. This computation is difficult though because in any given visual scene, there are many similar features, which create ambiguity for matching corresponding features registered by the two eyes. Modern computer vision algorithms solve this problem by sharing information from regions of high certainty about disparity with regions of low certainty about disparity. Using cross-correlation analysis, examining disparity tuning changes over time, and using a network model, we show that disparity-tuned neurons in the primary visual cortex (V1) can perform a similar process with organized recurrent connectivity. A more rigorous examination of the organization reveals that it matches predictions based on the statistics about spatial relationships found in 3D natural scenes. In regions where no features are present, disparity is impossible to estimate so evidence from other visual cues must be used to obtain a better estimate of depth. Due to shadows, the statistics between images and depth in 3D natural scenes predict a negative correlation between relative luminance and relative depth. We examined the relationship between relative luminance and binocular disparity tuning in V1 and find a similar negative correlation. With this correlation, near-tuned neurons will respond more to lighter surfaces and far-tuned neurons will respond more to darker surfaces, which is what is expected in the natural environment. The organization of the recurrent connectivity and the correlation between tuning in V1 both represent expected statistical trends found in 3D natural scenes and can be used by the visual system for inferring binocular disparity.