Can computational models of shape explain object perception?

SP Arun & RT Pramod

Centre for Neuroscience, Indian Institute of Science, Bangalore, India

ABSTRACT

Despite advances in computation and machine learning, computers are still far behind humans in vision. This is most likely because humans use a sophisticated object representation which is very different from that used in computers today. Another challenge is that object representations in computer vision and human vision have not been systematically compared on the same objects. To address this issue, we measured perceptual dissimilarity between objects in humans in a visual search (taking search difficulty as an index of target-distracter similarity). We then compared these observed dissimilarities against the dissimilarity predicted by a large number of state-of-the-art computational models of shape (e.g. Fourier descriptors, HMAX, Gabor filters, spatial pyramid etc.). In general, computational models were able to explain perceptual dissimilarity to a reasonable degree (r = 0.7-0.8 depending on the shape set). More interestingly, there were systematic deviations between all models and perceptual dissimilarity: for some pairs of objects, perceptual dissimilarity was greater than predicted by every model, whereas for other pairs, it was smaller. These systematic deviations are indicative of what is lacking in nearly all computational models of shape. Specifically, we propose that computational models of shape must incorporate some form of parts-based representation in order to account for the unexplained variation. We will also preview some related work (to be presented at the main VSS meeting) in which we have elucidated how object dissimilarity can be understood in terms of dissimilarities between parts.

METHODS & RESULTS

We measured perceived dissimilarity between pairs of animal and abstract shapes using visual search. The reciprocal of the search time was taken as a measure of perceived dissimilarity. We then used various Pixel-based, Boundary-based, Feature-based, Cortical Network based and Structural/Statistical models of object recognition to explain preceived dissimilarities. These computer vision models were able to predict the perceived dissimilarities between pairwise objects for many pairs. However, all models consistently under-estimated or over-estimated perceived dissimilarities for certain object pairs (Figure 1).

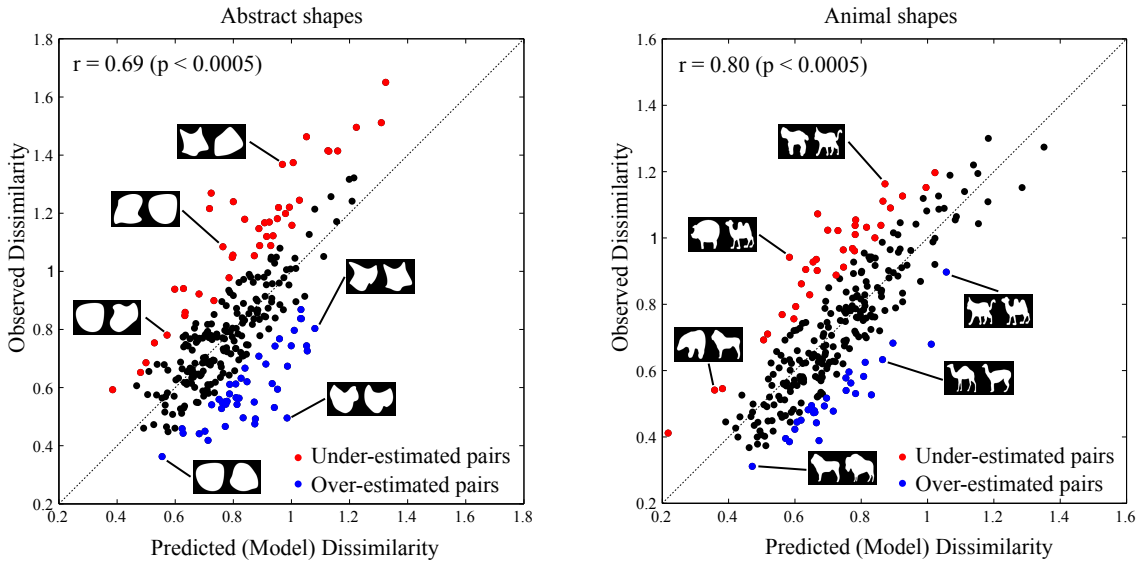

Figure 1 - Correlation between observed dissimilarities and dissimilarities predicted by a combination of all computational models, for Abstract shapes (*left*) and animal shapes (*right*). Object pairs that are consistently under estimated by all models are shown in *red* and those that are over-estimated by all models are shown in *blue*.