Classification of natural scenes using global image statistics

Jan Drewes[1], Felix Wichmann[2], Karl R. Gegenfurtner[1]
[1]University of Giessen, Giessen, Germany
[2]Max Planck Institute for Biological Cybernetics, Tübingen, Germany

The algorithmic classification of complex, natural scenes is generally considered a difficult task due to the large amount of information conveyed by natural images. Work by Simon Thorpe and colleagues showed that humans are capable of detecting animals within novel natural scenes with remarkable speed and accuracy. This suggests that the relevant information for classification can be extracted at comparatively limited computational cost.

One hypothesis is that global image statistics such as the amplitude spectrum could underly fast image classification (Johnson & Olshausen, Journal of Vision, 2003; Torralba & Oliva, Network: Comput. Neural Syst., 2003). We attempted to confirm this using linear discriminant analysis to classify a set of 11.000 images into animal and non-animal images. After applying a DFT to the image, we put the Fourier spectrum into bins (8 orientations with 6 frequency bands each). Using all of these 48 bins, classification performance on the Fourier spectrum reached 70%. However, performance was similar (67%) when only the high spatial frequency information was used and decreased steadily at lower spatial frequencies, reaching a minimum (50%) for the low spatial frequency information. Similar results were obtained when all bins were used on spatially filtered images. A detailed analysis of the classification weights showed that a relatively high level of performance (66%) could also be obtained when only 2 bins were used, namely the vertical and horizontal orientation at the highest spatial frequency band. In order to evaluate whether human subjects use the same information to do the task as our computer classifier, we performed several psychophysical experiments.

Performance of any classification mechanism - computer or human - depending on horizontal and vertical orientations should be severly impaired with images rotated by 90 degrees. We therefore measured reaction time and animal detection accuracy in a two alternative forced choice task, in which an animal image was presented next to a distractor image at one of five possible orientations (upright, 45, 90, 135 degrees rotated, or upside down). We found a modest effect of orientation on human classification performance: there was a 3% decrease in accuracy for obliquely oriented images that went along with a 10ms increase in reaction time. We then went on to test human classification performance with sets of animal and distractor images where the computer did particularly well or bad. Humans performed equally well on both sets of animal images, irrespective of how the computer did. However, humans were slightly better (5%) in correctly rejecting distractor images that the computer also classified well.

Our results show that in the absence of sophisticated machine learning techniques, animal detection in natural scenes is limited to rather modest levels of performance, far below those of human observers. If limiting oneself to global image statistics such as the DFT then mostly information at the highest spatial frequencies is useful for the task. This is analogous to the results obtained with human observers on filtered images (Kirchner et al, VSS 2004). We conclude that there is only little resemblance between the computer spectral classifier and human classification performance.