

Computer Vision vs. Human Vision¹

Bo Zhang

College of Information Science & Technology
Dept. of Computer Science and Technology, Tsinghua University
Tsinghua National Lab for Information Science & Technology
State Key Lab of Intelligent Technology and Systems
Beijing, China

Abstract

In object recognition (classification), it was known that the human brain processes visual information in semantic space mainly, that is, extracting the semantically meaningful features such as line-segments, boundaries, shape and so on. But by recent information processing techniques, these kinds of features cannot be detected by computers robustly so that in computer vision it's still difficult to process visual information as humans do. Computers have to process visual information in data space formed by the robustly detectable but less meaningful features such as colors, textures etc. Therefore, the processing methodology in computers is quite different from that in human beings.

In the talk, we will address the main principle of the image recognition (classification) approach in computer vision, its seedtime, main results and the difficulty faced recently. From digital cameras, there is a huge amount of 2D-image data. In computer object recognition (or classification), the data should be transformed into an object-invariant inner representation. In order to solve the problem, we need two key techniques, i.e., a robust detector and an invariant descriptor. People have attempted to solve the two key techniques for a long time but so far they didn't find any efficient solution. Human visual performances are still superior to that of computer vision greatly in many aspects. So as a future direction, computer vision should learn some things from neuroscience and brain science. We will discuss what computer vision can learn from human vision and how it will be affected by the new interdisciplinary research. We may still face many difficulties in the future.

Keywords: computer vision, human vision, object recognition, feature, detector, descriptor

About the Keynote Speaker



Professor Bo Zhang is a professor of Computer Science and Technology Department of Tsinghua University, the fellow of Chinese Academy of Sciences. In 1958 he graduated from Automatic Control Department of Tsinghua University, and became a faculty member since then. From 1980/02 to 1982/02 he visited University of Illinois at Urbana-Champaign, USA as a scholar. He is now the chairman of steering committee of Research Institute of Information Technology, Tsinghua University, the technical advisor of Fujian government, and the member of Technical Advisory Board of Microsoft Research Asia.

He is engaged in the research on artificial intelligence, artificial neural networks, genetic algorithms, intelligent robotics, pattern recognition and intelligent control. In these fields, he has published over 150 papers and 4 monographs, where 2 are English versions.

In the past 20 years, he presented the quotient space theory of problem solving, including the transformation, composition and reasoning methods among multi-granular spaces. Now the theory has become one of the main branches in granular computing. The theory has been applied to planning, search, machine learning, and information processing. Recently, he founded the center for neural and cognitive computation and the research group for multimedia information processing. The group has got some important results in image and video analysis and retrieval.

¹ Supported by the National Key Foundation R&D Project under Grant No. 2004CB318108, 2007CB311003