Which image processing algorithms best describe the minimal amount of visual information required for image recognition?

Cédric Laloyaux (claloyau@ulb.ac.be)
Cognitive Science Research Unit, Université Libre de Bruxelles
CP 122, Av. F.-D. Roosevelt, 50, 1050 Bruxelles, Belgium

Cédric Schmitt (cedric.schmitt@umontreal.ca)
LBUM, CHUM, Hôpital Notre-Dame, Pavillon J.A de Sève (Y-1619)
2099 Alexandre de Sève, Montréal, Québec, H2L 2W5, Canada

Introduction

In the framework of visual prosthesis, most methods use a video-camera in order to record visual information (Margalit et al., 2002). Determining the relevant level of information (e.g. contours, junctions) to be send to the retina, the optic nerve or a sensory substitution device for object recognition purpose, is a crucial aspect of raw images preprocessing (Boyle, 2002). For example, in Delbeke et al. (2002), a subject, with a four-contact optic nerve electrode, was able to discriminate simple patterns like "L" or "+" but nothing is said about the image processing algorithm to use in real-world conditions. In this research, we aim to investigate the minimal amount of information that different image processing algorithms provide, for object and scene recognition tasks.

Experiment

The goal of the study was to compare different image processing algorithms to determine which algorithms provide the best descriptions of the minimal amount of perceptual information required for image recognition.

Two categories of images were compared. The following image processing algorithms were applied to 15 images of objects (object category) and 15 pictures of indoor and outdoor scenes likely to be encountered by an observer in motion (mobility category): (1) Canny edge detector based on broadband frequencies (Canny BF), (2) Canny edge detector based only on Low frequencies (Canny LF), (3) The center on-off Marr's model, (4) a simple threshold method and (5) the Sobel (based on the directional gradient approximation of smooth image) (see Mallot, 2000).

Fifteen subjects performed 450 trials: 30 images X 3 image resolutions (32x32, 64x64 with a reliable, but slow, sub-sampling method and 64x64 with a poor, but fast, sub-sampling method) were processed by the five different algorithms. Subjects viewed a series of images: first a fixation cross for 1000 ms followed by a color image consisting of 480x480 pixels for 300 ms, and lastly two processed images: one being the original image processed (the target) and the other being any one of the other pictures from the same category (a distractor). In other words, an object target was presented with an object distracter. The target and the distracter were processed using the same image algorithm. The subjects' task was to press a key corresponding to the position of the target image on the screen. They were instructed to answer as quickly and accurately as possible.

Results and Discussion

ANOVA were performed on the error rates and the RT's. There was no trade-off effect. The RT's were significantly lower for the two 64x64 resolutions compared to the 32x32 [$F(2,28)=42.976$; $p<0.001$]. There was a main effect of photo type [$F(1,14)=50.247$; $p<0.001$], as RT were faster for objects than for scenes. This result supports the hypothesis that navigation based on artificial devices requires an adapted image processing step. There was also a main effect of image processing algorithms [$F(4,56)=6.6670$; $p<0.001$]. A planned comparison between the Canny and the Canny LF showed that RT's were significantly lower for the Canny LF ($p<0.05$). The thresholding method was also very efficient, and had the same level of efficiency as the Canny LF. However, the number of pixels to send after image processing was 10 times greater for the thresholding method than for the Canny LF. All together, the results showed that human performance differs greatly depending on the image processing algorithms used, and that these algorithms do not require the same amount of pixels in order for images to be minimally recognizable.

References

Boyle, J.R., Maeder, A.J., Boles, W.W. (2002). Image enhancement for electronic visual prostheses. *Australas Phys Eng Sci Med, 25*,(2), 81-6.

Delbeke, J., Wanet-Defalque, M.C., Gerard, B., Troosters, M., Michaux, G., Veraart, C. (2002). The microsystems based visual prosthesis for optic nerve stimulation. *Artificial Organs, 26*,(3), 232-234.

Margalit, E., Maia, M., Weiland, J.D., Greenberg, R.J., Fujii, G.Y., Torres, G., Piyathaisere, D.V., O'Hearn, T.M., Liu, W., Lazzi, G., Dagnelie, G., Scribner, D.A., de Juan. E., Humayun, M.S. (2002). Retinal prosthesis for the blind. *Survey of Ophatlmology, 47*,(4), 335-356.

Mallot, H.A. (2000). *Computational Vision,* Cambridge, MA: MIT press.