Cogn Process (2009) 10:209–210
DOI 10.1007/s10339-009-0270-2

EDITORIAL

Special corner: Visual Categorization and Image Management Systems

Valérie Bonnardel · Michael Oakes · John Tait

Received: 15 June 2009 / Accepted: 15 June 2009 / Published online: 1 July 2009

This special corner consists of three selected papers that were presented at the Visual Categorisation and Image Management Systems (VCIMS) workshop organised by V. Bonnardel, M. Oakes and J. Tait and sponsored by the Multimedia Knowledge Management Network chaired by S. Rueger that took place at the University of Sunderland on the 18 June 2006. The VCIMS workshop offered the opportunity to bring together Cognitive Psychologists, Visual Neuroscientists, Information Retrieval scientists to promote integration from interdisciplinary approaches on visual categorisation in natural and artificial systems.

Categorization is the ability to determine that an entity belongs to a particular group of objects and by this mean to recognize, differentiate and understand objects. Objects are classified in function of attributes defining their category memberships. There are different theoretical views that specify the nature of category-relevant attributes and the strategy in distinguishing between exemplars and non-exemplars of a given category. In the classical view, categories are mutually exclusive and collectively exhaustive discrete entities. They are characterized by a list of necessary and sufficient attributes, so that any object belongs univocally to one and only one category. More recent theories, such as prototype theory, acknowledge that natural categories are graded and that some members of the category, based on their properties, are more central than others. For these theories it is essential to evaluate the relative relevance of the different properties.

Hammer, Hertz, Hochstein and Weinshall propose a novel approach to determine the 'dimension weighting' (property relevance) in category learning. In their approach, learning strategies are based on a restriction indicating whether two exemplars belong (positive equivalence constraint, PEC) or not (negative equivalence constraint, NEC) to the same category. PEC differs from NEC on several aspects. PEC specifies within-category variations whereas NEC specifies between-category variations, and if NEC are more common than PEC (i.e., number of pair comparisons from between category objects is larger than that from within category objects) they are less informative. Indeed, if two objects are from the same category, it can be deduced that at least some of the shared dimensions are relevant to that category and all non-shared dimensions are irrelevant, but if two objects from different categories differ by more than one dimension, it is impossible to know which of these dimensions is relevant for discriminating between categories. Probably because of this difference in the amount of information, performances in a categorization task are better in PEC condition. When the amount of information is made equal in the PEC and NEC conditions, participants are divided in two groups: those who used the NEC strategy successfully and those who do not. These individual differences highlight fundamental distinctions between the two strategies. PEC is used more intuitively but not perfectly while NEC is potentially more accurate but not implemented by all participants. Interestingly, a similar asymmetry is observed in machine-learning where NEC is

V. Bonnardel (✉)
Department of Psychology, University of Winchester, Winchester, UK
e-mail: valerie.bonnardel@winchester.ac.uk

M. Oakes
Department of School of Computing and Technology, University of Sunderland, Sunderland, UK
e-mail: Michael.Oakes@sunderland.ac.uk

J. Tait
The Information Retrieval Facility, Opemgass 20b, 1040 Vienna, Austria
e-mail: john@johntait.net

not efficiently used because not adapted to algorithms based on prototype model of category learning.

A major issue for artificial vision systems in object recognition and categorization is the tolerance to transformations which is a central property of biological vision. This question is addressed in Rodrigues and du Buf model of visual categorisation. The model, biologically plausible, is grounded in the classical view of the functional architecture of the visual system based on the distinction of two parallel pathways extracting from the retinal input what is commonly called the 'what' and the 'where' information. The model is a multi-scale (i.e. different spatial scales) features extraction model, in which the low-level features (lines, edges and keypoints) extraction can be extended to higher-level processing to obtain object categorisation and invariance in translation, rotation and size. Lines, edges and keypoints are extracted in the primary visual cortex, multi-scale lines and edges representation are transferred to ventral pathways where 'what' information is processed, while the keypoint scale space provides the necessary information to construct saliency maps for the focus-of-attention and constitutes the basic information of the dorsal system ('where' information). In this integrated architecture, coarse-scale information propagates first and allows for rough categorisation, then finer scales are progressively performed under a top-down attentional control and used for identification. Thus defined, the model gives reasonable results for invariant object categorisation and recognition in particular in the domain of translations, rotations and scaling transformations.

Unlike in Psychology and Visual Sciences, the challenge for Information Retrieval systems is not to find the most realistic model or theory that accounts for human categorization, but to improve techniques to retrieve information. Optimisation of retrieval techniques has nowadays become an essential objective in situation where million of people access daily multimedia documents from Internet. Lin, Oakes and Tait paper addresses the question of improving retrieval techniques using Content-Based Image Retrieval (CBIR) in which applications are based on supervised (i.e. pairing input objects to desired outputs) machine-learning classifiers to match images and words. In a training phase, the system learns by being presented with examples of images and their relevant keyword classifications. Images are represented by a vector of features, which may describe such qualities as colour, texture or shape. This representation of images is a non-trivial task. One problematic situation occurs when an image consists of several colours, and is represented by an average colour value, rather than any of the colours which appear in the original image. Such noisy features cause some keywords never to be assigned to their correct images. In the paper by Lin et al., two techniques for useful representative feature selection are suggested to improve image classification. A Pixel Density filter (PDfilter) and Information Gain (IG) are proposed to solve the problem of spurious colour similarity. Their study first characterises each image by its predominant (as opposed to its average) colour value, then selects only the most suitable feature vectors as training data for classification of testing images. The results show that feature selection is a promising approach for large vocabulary keyword classification. Previous systems are able to distinguish just a few keyword categories, while the work described here is able to annotate images from a vocabulary of 190 concrete keywords.