Expert Systems with Applications 40 (2013) 6707–6712

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Reverse engineering expert visual observations: From fixations to the learning of spatial filters with a neural-gas algorithm

J.S. Cope [a,*], P. Remagnino [a], S. Mannan [a], K. Diaz [b], F.J. Ferri [b], P. Wilkin [c]

[a] Kingston University, London, UK
[b] Universitat de València, Valencia, Spain
[c] Royal Botanic Gardens, Kew, London, UK

ARTICLE INFO

Keywords:
Neural gas
Expert vision
Eye-tracking
Fixations

ABSTRACT

Human beings can become experts in performing specific vision tasks, for example, doctors analysing medical images, or botanists studying leaves. With sufficient knowledge and experience, people can become very efficient at such tasks. When attempting to perform these tasks with a machine vision system, it would be highly beneficial to be able to replicate the process which the expert undergoes. Advances in eye-tracking technology can provide data to allow us to discover the manner in which an expert studies an image. This paper presents a first step towards utilizing these data for computer vision purposes. A growing-neural-gas algorithm is used to learn a set of Gabor filters which give high responses to image regions which a human expert fixated on. These filters can then be used to identify regions in other images which are likely to be useful for a given vision task. The algorithm is evaluated by learning filters for locating specific areas of plant leaves.

1. Introduction

When viewing any detailed image, such as an advertisement, website or some particular object, the attention of the human visual system is attracted to certain features, known as salient regions. This process of observation is to a large extent innate and subconscious, although can become less so through prior knowledge of the observed image, or experience in viewing particular types of image. Research into eye-movement is involved in a patchwork of fields, including and beyond perceptual systems. The study of eye-fixation points and saccades (fast eye movements between points of interest/stimuli) can provide insight into cognitive processes such as written language comprehension, memory, mental imagery and decision making (Renniger, Verhese, & Coughlan, 2007). Eye movement research is of great interest in the study of neuroscience and psychiatry, as well as ergonomics, advertising and design (Wedel & Pieters, 2008). Since eye movements can be controlled, to some degree, voluntarily, and detected and recorded by modern technology with great speed and precision, they can now be used as a powerful input device for many practical applications in human–computer interactions (Richardson, Spivey, & Wnek, 2004).

Wearable eye-tracking devices allow collection of eye-movement information for natural scenes, involving the use of generally unconstrained eye, head, and hand movements. The most commonly sought eye-tracking metrics include the number, duration and location of fixations, both across the entire scene and within set areas of interest, and the sequence of movements between them, among many others (Jacob & Karn, 2003; Megaw & Richardson, 1979). Longer fixation periods generally indicate greater cognitive processing of the fixated region, possibly due to a higher level of detail or a lower scale feature of interest, and the percentage of total fixation dedicated to a particular area may indicate its saliency (Duchowski, 2007; Ryan, Duchowski, Vincent, & Battisto, 2010).

With sufficient knowledge and experience, an expert in a particular field can become highly efficient at analysing certain types of images. This could be a doctor searching for anomalies images produced by medical scanners, a botanist studying images of leaves to determine a plant's species, or a security personnel identifying suspicious behaviour in CCTV footage. Using advanced eye-tracking technology, we can capture and analyse in great depth the process through which a human expert analyses such images. This chiefly involves identifying their fixations, and analysing the sequence in which these fixations are visited. Through this it may be possible to an enable a computer system to accurately replicate the human expert's process. This could lead to advances in the use of computer vision techniques for performing such tasks, as it would allow more efficient processing of the images, and may reveal additional information which current techniques are overlooking.

In this paper we present a first step towards utilizing this type of eye-tracking data for computer vision purposes, concentrating

* Corresponding author. Tel.: +44 7549516762.
E-mail addresses: j.cope@kingston.ac.uk (J.S. Cope), p.remagnino@kingston.ac.uk (P. Remagnino), s.mannan@kingston.ac.uk (S. Mannan), katerine.diaz@uv.es (K. Diaz), francesc.ferri@uv.es (F.J. Ferri), p.wilkin@kew.org (P. Wilkin).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.05.042

on its use in the study of the classification of plant leaves, from the perspective of the expert in plant systematics, which uses tools based on morphology for identification and is one of the principal branches of study in plant biology. Plant systematists are responsible for the organisation and accessibility of plant diversity data which is underpinned by accurate identification and naming. Fig. 1 illustrates the typical sequences of fixations when an expert in plant systematics studies a leaf. In recent years there has been an increased interest in using computer vision to aid in these tasks (Andrade, Mayo, Kirkup, & Van Den Berg, 2008; Backes & Bruno, 2009; Gu, Du, & Wang, 2005). In our approach, we apply neural-gas algorithms (Martinetz & Shulten, 1991) for filter parameter learning, to discover a set of filter which are particularly well suited for identifying the fixation points on a image of a leaf.

The paper is structured as follows. Section 2 briefly outlines some of the previous work related to this paper. Our methodology is described in detail in Section 3. In Section 4 we discuss how our method will be evaluated and give the results of these experiments. Finally, in Section 5 we discuss what further work needs to be done to achieve our goal of expert vision replication.

2. Related work

In filter parameter learning (FPL) (Biem & Katagiri; Heidemann, 1996; Kurosawa, 2008), a set of image filters are described by some parameters whose values change through the course of some learning process. There have been numerous approaches to this problem. In Heidemann (1996), Heidemann presents an object recognition architecture based on feature extraction by Gabor filter kernels, and performs feature classification by an artificial neural network. The parameters of the Gabor filters are optimized to the specific problem by minimizing an energy function. These Gabor filters can then be used to extract features that can be more easily classified by a neural network. Alain and Shigeru in Biem and Katagiri (1994) used a discriminative feature extraction method to applied to a bank of filters for the modelling of speech. A method proposed by Kavukcuoglu, Ranzato, Fergus, and LeCun (2009) automatically learns the feature extractors in an unsupervised fashion by simultaneously learning the filters and the pooling units that combine multiple filter outputs together. The method generates topographic maps of similar filters that extract features of orientations, scales, and positions. By doing this, locally-invariant outputs are produced. In Gautama and Hulle (1999), force the filters to partition the input space in an equitable manner: each filter is tuned to a different frequency region and contributes equally to the extraction of localized features. Here we learn a set of Gabor filters for processing images, due to their well known properties in extraction of features from their parameters of frequencies, orientations, and smoothing of the Gaussian envelope (Grigorescu, Petkov, & Kruizinga, 2002; Randen & Husoy, 1999; Li, Mao, Zhang, & Chai, 2010; Chi, Houqiang, & Chao, 2003). Furthermore, links have been identified between Gabor filters and the human visual system (Daugman, 1985), and as such they may have added benefit for our purposes.

In the field of neural networks many different architectures and training rules exist, from the perceptrons (from single-unit to multilayer versions), Hopfield-type recurrent networks (including probabilistic versions strongly related to statistical physics and Gibbs distributions) and the self organizing map (SOM), among others (Feldkamp, 1996). In a self-organising map, the network being trained has a fixed topology throughout, however there exist several variants where, based on errors within the network, elements of the network are added or removed. The neural-gas algorithm (Martinetz & Shulten, 1991) is one such variant, which uses a fixed number of nodes, and adds and removes connections so that for every input pattern, the two closest nodes are connected in the final network. In short, the organization of neurons, according to their distance to the input pattern, and subsequent modification of its reference vector, produces the expansion of them within the input space. Subsequently, by adding and deleting of the edges a triangulation between different processing elements is provides. An extension to this, the growing-neural-gas algorithm (GNG) (Fritzke, 1995) is initialised with just two nodes, and adds more over time. Furthermore, it removes any nodes which have become separated from the network in an unused area of the space. This removes the requirement for *a prior* knowledge about the topological dimension of the space of input vectors (Mendona Ernesto Rego, Araujo, & de Lima Neto, 2010). We have based our approach on this form of the neural-gas algorithm.

3. Methodology

Our approach here is find a set of image filters that can be used to efficiently identify possible fixation points on an image of a plant leaf. Firstly, we collected data on where such fixation points lie by using an eye-tracking device to capture a botanist's eye movements as he studies a series a leaf images. Each leaf was shown to the botanist for a set period of time, during which he was asked to verbally give as much information as possible about the leaf. This ensured the manner in which the leaves were studied was realistic and relevant. We use the fixations which have been discovered as input into an algorithm which attempts to find a set of filters which give high responses to fixation windows (Section 3.1). The filters we learn are based upon the Gabor model (Section 3.2). The learning is performed using a variant of the growing-neural-gas algorithm (Section 3.3).

3.1. Fixations and filter responses

We define a fixation point as being a point on an image where a person focuses their attention for some amount of time (for

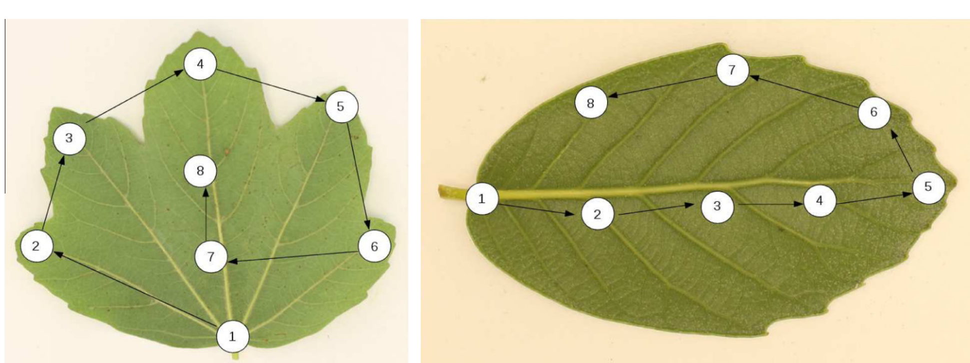

Fig. 1. Examples of typical sequences of fixation for eye-tracker data from an expert botanist.