Available online at www.sciencedirect.com
ScienceDirect
Neural Networks 20 (2007) 260–273
Neural Networks
www.elsevier.com/locate/neunet

Component-based visual clustering using the self-organizing map

Mustaq Hussain*,[1], John P. Eakins[2]

School of Informatics, University of Northumbria at Newcastle, NE1 8ST, United Kingdom

Received 18 November 2003; accepted 24 October 2006

Abstract

In this paper we present a new method for visual clustering of multi-component images such as trademarks, using the topological properties of the self-organizing map, and show how it can be used for similarity retrieval from a database. The method involves two stages: firstly, the construction of a 2D map based on features extracted from image components, and secondly the derivation of a *Component Similarity Vector* from a query image, which is used in turn to derive a 2D map of retrieved images. The retrieval effectiveness of this novel component-based shape matching approach has been evaluated on a set of over 10 000 trademark images, using a spatially-based precision–recall measure. Our results suggest that our component-based matching technique performs markedly better than matching using whole-image clustering, and is relatively insensitive to changes in input parameters such as network size.

Keywords: Component matching; Content-based image retrieval; Feature extraction; Image segmentation; Self-organizing map; Similarity measures; Trademark image retrieval

1. Introduction

The number and variety of image collections available in electronic form has risen rapidly over recent years, leading to both opportunities and problems for image users. The difficulties involved in finding a desired image in a large collection has led to increasing interest in automatic techniques for content-based image retrieval (CBIR). Most CBIR techniques (see Smeulders, Worring, Santini, Gupta, and Jain (2000) for a comprehensive review) operate by computing similarity measures between stored and query images from the values of automatically extracted features. These typically describe visual characteristics of the image such as colour, texture and shape. To date, no single technique has been developed that can accurately describe a general image. Hence researchers have tried to develop a community of different models (Picard, 1996) to describe different aspects of image appearances.

One such approach has been retrieval by shape similarity, but this has proved particularly challenging. Over the last decade, researchers have proposed a rich variety of techniques, including comparison of boundary segment chains (Mehrotra & Gary, 1995), elastic deformation of templates (Pentland, Picard, & Sclaroff, 1996), Fourier descriptors (Zahn & Roskies, 1972), moment invariants (Hu, 1962), Zernike moments (Teh & Chin, 1988), edge direction histograms (Jain & Vailaya, 1996), the angular radial transformation (Sikora, 2001) and wavelets (Mallat, 1989). However, empirical tests of retrieval effectiveness (e.g. Faloutsos et al. (1994)) suggest that the problem is far from solved.

Despite their early promise, neural networks have not been widely used for large-scale image retrieval applications (Oja, Laaksonen, Koskela, & Brandt, 1999). The Self-Organizing Map (SOM) or Kohonen network (Kohonen, 2001), however, has been used as the basis for the PicSOM system (Koskela, Laaksonen, Laakso, & Oja, 2000), an image retrieval system using multiple Tree Structured SOMs (Koikkalainen & Oja, 1990). The tree structured SOM (TS-SOM) is a pyramid structured SOM that progressively gets larger as one descends the hierarchy. Each TS-SOM is trained on one particular feature – colour, texture, edge direction and Fourier descriptors – and

* Corresponding address: 22 Kingsley Place, Newcastle-upon-Tyne, NE6 5AN, United Kingdom. Tel.: +44 0 191 276 2119; fax: +44 0 191 225 5933.
E-mail addresses: mustaq.hussain@unn.ac.uk, mustaq.hussain@atosorigin.com (M. Hussain), eakins@cs.york.ac.uk (J.P. Eakins).
[1] Present address: Atos Origin, United Kingdom.
[2] Department of Computer Science, University of York, YO10 5DD, United Kingdom.

0893-6080/$ - see front matter
doi:10.1016/j.neunet.2006.10.004

Symbols

$\#\{\}$	Cardinality
$\delta()$	Unit step function
exp	Exponential
$\mathbf{C}^A$	Component Similarity Vector for image $\mathbf{T}^A$
$\mathbf{C}_{c=0}$	Set of CSVs with all elements zero
c	Node of best matching unit
c_i^A	ith element of CSV for image $\mathbf{T}^A$
$f_{i,j}^A$	Feature measure j for component i of image $\mathbf{T}^A$
$\mathbf{I}_0$	Genuine trademark
$\mathbf{I}_i$	Image identified by indexers as being similar to $\mathbf{I}_0$
m	Number of components making up query image
$\mathbf{N}_i$	Neighbourhood of image component i
n	Number of components making up test image; number of query sets
$P(r, \mathbf{Q})$	Radial precision to distance r for image $\mathbf{Q}$
$P_{\text{avg}}(r)$	Average precision to distance r for all query sets
$\mathbf{Q}$	Query image
$\mathbf{Q}_\mathbf{S}$	Set of queries
$\mathbf{q}_i$	Component i of query $\mathbf{Q}$
$\mathbf{r}_i$	Positional vector of node i on SOM map
r	Neighbourhood radius
$\mathbf{T}^A$	Test image A
$\mathbf{t}_i^A$	Component i of image $\mathbf{T}^A$
$\mathbf{t}^A$	Collection of components making up image $\mathbf{T}^A$
$\mathbf{S}$	Expected query result set
$S_\#(\mathbf{T}^A, \mathbf{Q})$	Component hit count similarity measure
$S_n(\mathbf{T}^A, \mathbf{Q})$	Normalised best matching similarity measure
$R(r, \mathbf{Q})$	Radial recall to distance r for image $\mathbf{Q}$
$R_{\text{avg}}(r)$	Average recall to distance r for all query sets

outputs from multiple SOMs are combined to retrieve images. Other SOM-based image retrieval systems include those of Breiteneder, Merkl, and Eidenberger (1999), who describe a coat-of-arms retrieval system, Barbalho, Neto, Costa, and Netto (2001), whose system uses a compressed image vector to store and retrieve images from a hierarchical SOM, and Garcia-Berro, Torres, and Isern (2003), who use a SOM to identify potential white dwarf stars.

Research into trademark image retrieval has become increasingly active over the last few years (Eakins, 2001, Chap. 13). This is because trademark images provide a good test-bed for techniques in shape retrieval algorithms (colour and texture are rarely important in trademark image retrieval) and because of their economic importance to business. It is, however, a very demanding application in one respect. In contrast to application areas such as fashion or journalism, where it is normally sufficient to retrieve *some* images meeting the user's need, the nature of the trademark registration process requires that *all* potentially relevant images be retrieved. No system yet described in the literature is capable of meeting this exacting requirement.

Currently there are two main approaches to trademark image retrieval: one group of researchers extracts and compares features from trademark images *taken as a whole*; the other regards trademark images as a set of discrete components which are best matched on a component-by-component basis. The earliest example of the first approach was the TRADEMARK system (Kato, 1992). This mapped normalized trademark images to an 8×8 pixel grid, and calculated a *GF-vector* for each image from various pixel frequency distributions, allowing query and stored images to be matched by comparing GF-vectors. Other researchers following this approach have included Jain and Vailaya (1998), who use a combination of edge direction histograms and moment invariants, Kim and Kim (1998), who use Zernike moments, and Ravela and Manmatha (1999), who use histograms of local curvature and phase.

The second approach is best exemplified by the STAR system developed by Wu, Lam, Mehtre, Gao, and Narasimhalu (1996), which allows human indexers to segment trademark images into perceptually meaningful components, from which shape features such as Fourier descriptors and moment invariants are extracted. Overall similarity between trademarks is expressed as a distance measure computed from the weighted sum of component distances. Peng and Chen (1997) take this principle one stage further, approximating each image component as a set of (possibly overlapping) closed contours, and matching images in a hierarchical fashion. Our own ARTISAN[3] system (Eakins, Boardman, & Graham, 1998) is also based on similar principles, using multi-level matching based on simple global features calculated both from individual image components and from perceptually significant families of components. More recent versions of the system (Eakins, Edwards, Riley, & Rosin, 2001) incorporate multiresolution analysis to remove texture and group low-level components into higher-level regions, as well as a wider range of shape and structural features.

Which of these two approaches is more effective has not yet been conclusively established. However, comparative studies suggest that component-based matching is capable of achieving significantly higher retrieval accuracy than whole-image matching (Eakins, Riley, & Edwards, 2003). It also is inherently more flexible, in that it can also support part-image matching. Against this, it is more computationally expensive, and has to rely on accurate image segmentation, a far from trivial problem.

Another important issue for image retrieval systems is how results are presented. In most trademark image retrieval systems results are presented as an ordered list, according to some measure of similarity. However, these 1D lists can make it difficult to see how similar non-adjacent images are related. One way to address this is to place images on a 2D surface where their 2D positioning can reflect the mutual distances between images, in some feature space. This type of visualisation can provide visual clues as to why particular trademarks cluster around the query, and why others have been placed further away, in an effective and intuitive way for a user,

[3] *Automatic Retrieval of Trademark Images by Shape Analysis.*