

## CONTENT-BASED PROFILING OF IMAGE COLLECTIONS: A SOM-BASED APPROACH

D. Deng

Department of Information Science, University of Otago, New Zealand  
[ddeng@infoscience.otago.ac.nz](mailto:ddeng@infoscience.otago.ac.nz)

### ABSTRACT

Content-based image retrieval techniques have been under intensively research, focusing on extracting effective low level visual features for indexing and enabling fast and accurate retrieval of individual images by matching the feature indexes. In this paper we propose to extend the content-based approach towards the problem of multimedia collection profiling and comparison. Our approach is to carry out visual feature clustering using self-organizing maps, and then to apply distance measures on the generated feature maps for similarity evaluation. A modified Hausdorff distance is defined over the feature maps and further verified in an experiment using four image collections. Some preliminary results are presented with a comparison of different distance measures obtained from profiles generated by two feature schemes.

**Keywords:** Content-based image retrieval, Self-organizing maps, Point sets.

### 1. INTRODUCTION

Over the last decade, content-based image and video indexing and retrieval has been a popular research topic, much owing to the ever increasing use of multimedia in personal entertainment, education, commercial websites, and personal mobile communications. Despite some breakthroughs made in the field, it is generally understood that the problem is still far from being solved (Smeulders et al., 2000). Aimed at effective multimedia asset management and efficient information retrieval, a typical content-based image retrieval system (e.g. Smith et al., 1996, Carson et al., 2002) works basically on low-level visual features such as colour, texture, shape or regions. Owing to obstacles in object recognition and image understanding, it is still hard to link these low level features to high-level concepts that correspond to objects and their semantic information within the image. Similarity defined in the low-level feature space cannot reflect faithfully the semantic similarity of different images. The similarity of image contents can vary among different levels - locally or globally, on different characteristics, or on account of different psychological effects. It is truly in the eyes of the beholders and rather hard to define explicitly. Two images can have the same or similar low-level visual attributes, e.g. the same colour map, while containing totally different semantic contents. Consequently image search engines working on low-level features alone may sometimes produce a great deal of nonsense results (Santini et al., 1998), even though techniques such as spatial colour features (Huang et al., 1998), joint histograms (Pass and Zabih, 1999), image classification (Hirata and Mukherjee, 2000) and conceptual description (Mukerjee et al., 1999) have been investigated to more or less improve the retrieval quality. The semantic gap between content-based representation and semantic description of images and videos is still hard to overcome.

On the other hand, with the lack of semantic description in multimedia storage nowadays, content-based representation and retrieval as a rudimentary solution can still provide useful interfaces for browsing and search activities. In content-based image retrieval, generally one is interested in searching out interested patterns or images. There are however circumstances where there is a need to compare different image collections - e.g. images stored in folders named 'family photos' in different locations, or two collections of architecture photos. People may want an easy way to compare their music tastes by profiling their MP3 collections and finding how much the profiles differ from each other. A multimedia collection

management tool should be equipped with such kind of capabilities. In this paper we propose to make use of content-based indexing and retrieval techniques to tackle the problem of collection profiling and comparison. A neural network model is employed for self-organised profiling of image collections. The neural structure of the profiles also provides a graphical interface for collection navigation and visual comparison. To establish quantitative assessment of the similarity of the profiles, a number of dissimilarity measures based on point set distances are implemented. Among them it is found a modified soft distance measure performs favourably in a case study problem.

The rest of this paper is organised as follows. In Section 2 we present a brief introduction on the computational model, including the profiling of collections via self-organizing feature maps, and different distance measures defined on the feature maps. In Section 3 four image collections are tested with our computational model and results from different measures are compared. Finally the paper is concluded in Section 4 with a discussion on some future works.

## 2 THE COMPUTATIONAL MODEL

### 2.1 Self-organizing maps

The content-based profiling problem can be regarded as to extract representative prototypes from the feature space generated from the original image collections. Therefore, a number of clustering or vector quantisation algorithms can be employed for this purpose. Among various clustering algorithms available, one neural network model of special interest to us is the Self-Organizing Map (SOM) (Kohonen, 1997). It has been applied widely in a number of information retrieval systems, such as in SOMLib (Raubert and Merkl, 1999) and PicSOM (Laaksonen et al., 1999).

SOM features in carrying out vector quantization and multi-dimensional scaling at the same time. The map, usually set in 2-D or 3-D topology, consists of a regular lattice of neurons set in hexagonal or rectangular topology. Each neuron is associated with a weight vector. The map attempts to perform localised clustering on these node vectors, while in the meantime the ordering on the lattice works to match similar inputs to the same node or nodes close to each other, and dissimilar inputs onto nodes far from each other. The nodes are sometimes also called as units, and unit vectors as prototypes.

Assume we have a  $N$ -prototype SOM to train. Denote  $w_i(t)$  as the weight vector associated with node  $i$ . Given an input  $x(t)$ , the algorithm first finds the best-matching unit (BMU)  $w_b$  among all prototypes, i.e.

$$b = \arg \min_i \|x(t) - w_i(t)\|, \quad i = 1, \dots, N \quad (1)$$

The weight vectors are then updated according to the following learning rule:

$$w_i(t+1) = w_i(t) + \gamma(t)h_{b,i}(t)[x(t) - w_i(t)] \quad (2)$$

where  $h_{b,i}(t)$  is a neighbourhood function centred at BMU and shrinking over time, and  $\gamma(t)$  the learning rate decreasing over time. There are some variants proposed to this original learning rule, but generally it has been shown that these learning rules lead to the convergence of unit vectors.

### 2.2 Collection profiling by SOMs

SOM has a number of traits that make it a natural choice for collection profiling:

- Good visualisation ability. Network nodes are located on a low-dimensional lattice easy for visualisation and human interpretation. This also keeps users' navigation in the high-dimensional feature space traceable on a low-dimensional map.
- Topology preserving. Similar inputs are mapped onto the same node or nodes in a neighbourhood on the map. This means that similar images can be closely mapped onto the grid, also making browsing easier and robust. Hierarchical implementation of the maps is also made possible.
- Density matching. Although not being able to match exactly to the probability distribution underlying in the input data (see Haykin 1999, page 460-461), the algorithm of SOM manages to represent a cluster of frequently occurring input stimuli by a larger area in the feature map. If we denote the number of nodes in a small volume  $dx$  over the input space  $X$  as  $m(x)$ , it is proved that SOM manages to achieve  $m(x) \propto p^{2/3}(x)$ , where  $p(x)$  is the probability density function of the input  $x$  (Ritter, 1991).

For these reasons it is not surprising that a number of applications have been built on the SOMs or tree-structured SOMs for image retrieval. To extend the use of SOM for collection profiling is as straightforward as using SOM trained on content-based visual features to construct a snapshot of the whole collection so that operations such as browsing, search, and comparison can be carried out in efficiency. For purposes of navigation and browsing hierarchical structures are often employed, but for profiling oriented for quick comparison across collections, we adopt flat SOMs to facilitate efficient calculation. On the other hand, little mathematical work has been done on the optimal structure design, topology preserving and density matching ability etc. for hierarchical feature maps.

### 2.3 Comparison of profiles

Although the SOM algorithm has been used widely for data analysis in all kind of applications, the problem of comparing two different feature maps has received little treatment in the literature. A visual approach is to project all feature maps in a graph of low dimensionality, using principal component analysis (PCA) for instance. While this can help to the visual exploration of multimedia collections, it gives little quantitative information about their similarity. In (Kaski and Lagus, 1996), a dissimilarity measure is proposed based on the evaluation of the goodness of SOMs by comparing the shortest paths on the maps of a given pair of data samples. To calculate the distance measure all pairs of data samples need to be matched onto the SOMs in comparison, and this can be rather time consuming. This method was used for comparison of word category maps generated by SOM in (Honkela, 1997). When dealing with high dimensional feature maps generated from a large volume of multimedia collections, the efficiency of such an approach will however be in question, as the plausibility of retrieving the whole data set for calculation can hardly be assumed. On the other hand, given that those feature maps have formed good profiles of the original data collections, a comparison process based on the map prototypes rather than the original data sets would be much more efficient.

As SOM basically is a clustering algorithm that extract a few prototypes from the overall joint feature set, we start by viewing the SOM comparison problem as measuring the distance between two point sets, which leads to a number of point set distance measures proposed in the context of computational geometry, image processing, and science of philosophy.

#### 2.3.1 Hausdorff distance

Given two point sets  $X$  and  $Y$ , the Hausdorff distance from  $X$  to  $Y$  is defined by

$$h(X, Y) = \sup_{x \in X} \inf_{y \in Y} d(x, y) \quad (3)$$

where  $d$  is a  $L_p$  metric, but usually the Euclidean distance is used.

The Hausdorff metric between sets  $X$  and  $Y$  is define as

$$HD(X, Y) = \max\{h(X, Y), h(Y, X)\} \quad (4)$$

It is found that the Hausdorff distance is very sensitive to outliers in the point sets. Some modification can be done, for example, by generalising the maximum with a quantile or a median. The Hausdorff distance has been applied in fractal image compression, shape matching and object detection etc.

#### 2.3.2 Earth Mover's Distance

The Earth Mover's Distance (EMD) (Rubner et al., 1998) is defined over weighted point sets. Suppose each point set is configured by a normalised weight set. Denote a point set as  $A = \{a_1, a_2, \dots, a_m\}$ , with  $a_i = \{(x_i, w_i)\}$ ,  $x_i \in R^k$ , and  $w_i \in R^+ \cup \{0\}$ . EMD calculates the minimum amount of work needed to transform one configuration to another by moving weight under constraints.

Denote the set of all feasible flows as  $F = \{f_{ij}\}$ , where  $i$  is a point label for set  $A$ , and  $j$  for  $B$ . The following relations should hold:

1.  $f_{ij} \geq 0, i = 1, \dots, m, j = 1, \dots, n$
2.  $\sum_{j=1}^n f_{ij} \leq w_i, i = 1, \dots, m$
3.  $\sum_{i=1}^m f_{ij} \leq u_j, j = 1, \dots, n$
4.  $\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min(W, U)$

Here  $W$  and  $U$  are the total weights of  $A$  and  $B$  respectively. These constraints ensure that, for instance, each flow of weight is non-negative; a point at the 'sender' can not send more weight than it holds, and a point at the 'receiver' does not receive more weight than it needs.

The EMD between the two point sets can then be define by

$$EMD(A, B) = \min_{f \in F} \sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij} \quad (5)$$

EMD has been applied in image retrieval for similarity comparison of global colour features, texture features, and shapes.

To make EMD eligible for SOM comparison, weights need to be assigned to the prototype nodes. An easy solution is to map the original data set onto the trained SOM and assign the probability of each node being selected as the BMU onto the node as the weight. Due to the probability density matching characteristic of SOM, it is however likely for the nodes to share a flat firing rate distribution. Also to acquire the firing rate over the entire data population can be rather time-consuming, unless on-line resource information is tracked as in some variants of the algorithm (e.g., Fritzke, 1995). In practice we assign a uniform weight to all nodes in a map for sake of efficiency.

### 2.3.3 Sum of Minimum Distances

In (Eiter and Mannila, 1997) the sum of minimum distances (SMD) as a similarity measure is discussed. It is defined as:

$$SMD(X, Y) = \frac{1}{2} \left( \sum_{x \in X} \min_{y \in Y} d(x, y) + \sum_{y \in Y} \min_{x \in X} d(x, y) \right) \quad (6)$$

SMD was proposed in the domain of philosophy of science and to our knowledge it has hardly been used for data analysis.

### 2.3.4 Sum of Average Neighbour Distance (SAND)

The calculation of sum of minimum distance between two SOMs is straightforward. However, it is noted that the measure ignores the established structure between prototypes within a SOM. As stated earlier a learned SOM not only quantises its high dimensional feature space but also self-organises into a grid structure that reflects the probability distribution of the feature data. The neighbourhood topology within a SOM can be examined to further characterise the original feature space. Taking into account of this, we propose a modified sum of minimum distance, so called *sum of average neighbour distance* (SAND).

The same as in the calculation of HD and SMD, for a prototype  $x \in X$ , its BMU, i.e., the prototype  $y_b \in Y$  with the minimum distance to  $x$ , is found. To calculate SAND, this minimum distance is averaged with the distances between  $x$  and the neighbours of  $y_b$ , before it is summed across all population of  $X$ . The same process is then repeated for set  $Y$ .

The calculation process can be summarised in the following steps:

1. Find the BMU  $y_b \in Y$  for any  $x \in X$ , with

$$b = \arg_{y \in Y} \min \|x - y\| \quad (7)$$

2. Find out all best-matching pairs  $(\alpha, \beta)$  between the neighbourhood of  $x$  and  $y_b$ , and calculate the averaging distance:

$$d_n(x) = E \{ \|\alpha - \beta\|, \alpha \in \Omega(x), \beta \in \Omega(y_b) \} \quad (8)$$

Here  $\Omega(\cdot)$  denotes the neighbourhood of a map node.

3. Sum up the individual measures:

$$SAND(X, Y) = \frac{1}{2} \left( \sum_{x \in X} d_n(x) + \sum_{y \in Y} d_n(y) \right) \quad (9)$$

The rationale behind this scheme is the probability density matching ability of SOM. By examining the matching among a map neighbourhood can tell the difference between maps of similar range of spatial span but originated from different probability distributions. As density differing in the original feature space will result in, on the low dimensional maps, either dense grids or sparse grids, their difference can be reflected by SAND better than a plain point-to-point measure.

### 3 Experiments and results

We use pictures downloaded from the SUNET FTP site to experiment on our collection profiling and comparison methods. There are four categories of images used, namely *views*, *sports*, *animals*, and *vehicles*. Each category holds images differing in background or presenting different objects in front. There are 413 images in ‘animals’, 170 in ‘views’, 391 in ‘sports’ and 356 in ‘vehicles’. Selected thumbnails of each category are shown in Fig.1 (a)-(d).



Figure 1: Thumbnails from the collections: (a) ‘views’, (b) ‘vehicles’, (c) ‘sports’, and (d) ‘animals’.

The first feature scheme we explore is a simple one, using regional average colours (RAC) in five non-overlapping zones, the same as in PicSom (Laaksonen et al., 1999). For sake of simplicity, feature sets of all four categories are clustered on  $8 \times 8$  SOMs. These SOMs can then be used in a browsing interface that facilitates navigation of images guided by their similarity in colour distribution.

On the other hand, we apply our proposed map comparison methods to measure their distances, so as to estimate the closeness of contents of different image collections. Another feature for texture analysis is employed, using a set of Gabor filters in four frequency levels and eight orientations. The energy histogram of each image is transformed into a 32-dimension feature vector.

Fig.2 shows the RAC profile generated as the feature map obtained from the ‘views’ collection.

For visual comparison of four profiles generated on the RAC feature scheme, they are all displayed in a 2-D plot generated by the projection of their prototypes over the first two eigenvectors of the overall prototype set. This is shown in Fig.3. Comparing the point ‘clouds’ in Fig.3, it gives visual clues to the closeness of the image collections. In this case, taking the ‘sports’ map as the base, we may say the closeness of the collections can be ranked as ‘animals’:1, ‘views’:2, ‘vehicles’:3, although the latter two are almost of the same offset.

Although it is usually the case that visual comparison provides no quantitative information and is hence not accurate, nevertheless we use it in this study as a reference to evaluate the validity of different distance measures.



Figure 2: The profile of image collection ‘views’ generated using the RAC feature.

In Table1 all distance measures calculated between the ‘sports’ collection and the other three using the RAC feature are listed together with their CPU time spent. It is noted that Hausdorff, EMD, and SAND all result in similar ranking of the similarity between collections - ‘animals’ the first, ‘views’ the second, and ‘vehicles’ the third, rather close to the previous visual assessment. Sum Min fails to achieve such a ranking. On the other hand, while most measures are efficient to calculate, EMD requires the longest time, requiring more than 0.20 seconds to complete while others need only about 0.01 seconds. This result agrees with other empirical studies made on EMD used for image dissimilarity computation (Puzicha et al., 1999). The CPU time is collected from a Linux 2.2 system running on a Pentium-II PC. Although the difference is not significant here, for applications with larger maps to compare it is a factor to consider.

Table 1: Distance calculated between the RAC profiles of ‘sports’ and other three collections.

Distance Measures	Collections			CPU Time
	Animals	Views	Vehicles	
HD	125.4	153.3	135.1	0.01
EMD	73.4	101.0	102.9	0.10
SMD	55.4	66.1	56.7	0.01
SAND	89.4	118.6	109.0	0.02

Table2 gives the results with profiles generated by the GFH feature. Since the feature codes are in a different range the distance values are also quite different when compared with those in Tab.1. It is easy to see, however, that the similarity ranking of the collections is similar.

Table 2: Distance calculated between the GFH profiles.

Distance Measures	Collections			CPU Time
	Animals	Views	Vehicles	
HD	0.54	0.64	0.59	0.01
EMD	0.23	0.29	0.59	0.12
SMD	0.33	0.38	0.26	0.02
SAND	0.35	0.39	0.31	0.04

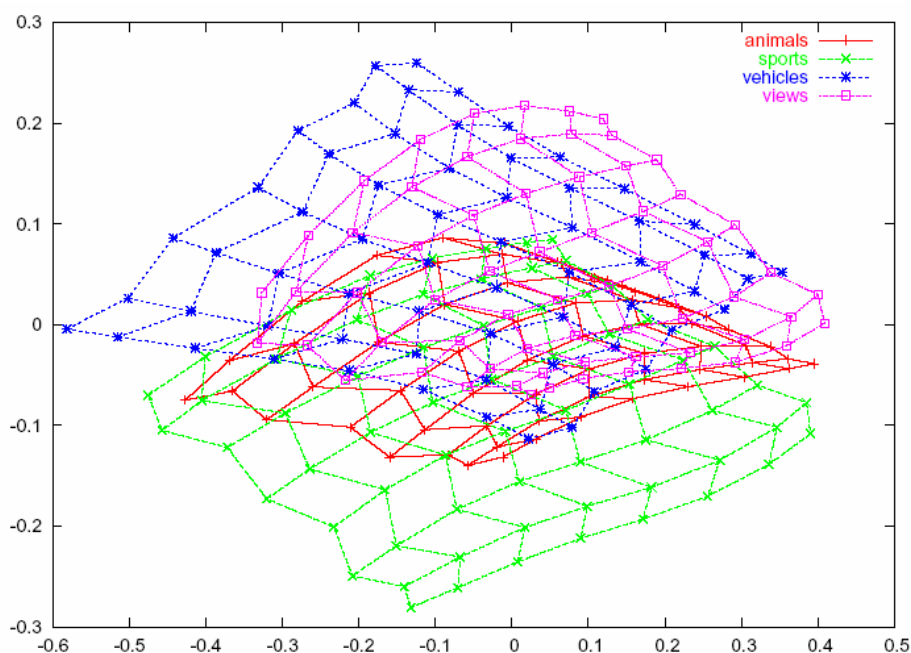


Figure 3: Four RAC-feature SOMs visualised using PCA.

#### 4 Conclusions

We propose to use self-organizing maps trained on low level content-based features extracted from image collections for content-based profiling and comparison. A number of different distance measures are examined in a four-category problem with more than 1000 images in total. We have found a modified measure, sum of average neighbour distance, taking account of existing topology in a SOM, gives comparable accuracy in similarity measure and works in better efficiency compared to the Earth Mover's Distance. We would like to test our methods on image collections of larger scale with more complicated feature schemes. On the other hand, a SOM-based feature clustering process also presents a good visual data mining interface and it will be a future direction for us to explore the feature space and establish some connections between feature clusters and semantic objects, which we hope may eventually help to solve the semantic gap problem.

While the SOM seems a natural and effective approach for document and multimedia data analysis, a few drawbacks exist as well. There is no efficient mechanism for a feature map, once trained, to adapt its own size when there is a need to allocate new resource for novel inputs. The lack of incremental learning ability in the SOM also makes on-line adapting of the network implausible. Considering the use of profiling for multimedia collections under constant variation, it is desirable to adapt the existing profiles with new data without the retraining of the whole model. This is hard to achieve with the SOM.

Apart from the SOM, other self-organising neural networks are being considered, such as GNG (Fritzke, 1995) and ESOM (Deng and Kasabov, 2003). While these models have improved on-line learning ability, they have to deal with the difficulty in visualisation once freeing their network structure from topology constraints. Also new distance measures may need to be developed for profiles generated by the new approaches. By employing better on-line learning ability into our computational model, and investigating the use of multiple features in profiling and comparison, we look forward to extend this approach for audio and video clips.

**Acknowledgment** This work is supported by Otago Research Grant ORG200200621 and partly by FRST AITX0201, Foundation for Research Science and Technology, New Zealand.

## References

Carson, C., Belongie, S., Greenspan, H., and Malik, J. (2002) Blobworld: A system for region-based image indexing and retrieval, *IEEE Transaction on PAMI*, **24**(8), 509–516.

Deng, D. and Kasabov, N. (2003) On-line pattern analysis by evolving self-organizing maps, *Neurocomputing*, **51**, 87–103.

Eiter, T. and Mannila, H. (1997) Distance measures for point sets and their computation, *Acta Informica*, **34**(2), 109–133.

Fritzke, B. (1995) A growing neural gas network learns topologies, in Tesauro G., Touretzky D., and Leen, T. (eds.) *Advances in Neural Information Processing Systems*, **7**, 625–632.

Haykin, S. (1999) *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 2nd edition.

Hirata, K. and Mukherjea, S. (2000), Integration of image matching and classification for multimedia navigation, *Multimedia Tools and Application*, **11**, 295–309.

Honkela, T. (1997) Comparisons of self-organized word category maps, in *Proceedings of Workshop on Self-Organizing Maps*, Espoo, Finland, 298–303.

Huang, J., Kumar, S., Mitra, M. and Zhu, W. (1998) Spatial color indexing and applications, in *Proceedings of the Sixth International Conference on Computer Vision (ICCV'98)*, IEEE Computer Society, Washington, DC, 602–607.

Kaski, S. and Lagus, K. (1996) Comparing self-organizing maps, in Vorbruggen, J. and Sendhoff, B. (eds.), *Lecture Notes in Computer Science*, **1112**, Springer, Berlin, 809–814.

Kohonen, T. (1997) *Self-organizing Maps*, Springer-Verlag, second edition.

Laaksonen, J., Koskela, M. and Oja, E. (1999) Content-based image retrieval using self-organizing maps, In *Proceedings of Third International Conference on Visual Information Systems (Visual'99)*, Amsterdam, The Netherlands, 541–548.

Mukerjee, A., Gupta, K., Nautiyal, S., Singh, M.P. and Mishra N. (1999) Conceptual description of visual scenes from linguistic models, *Image and Vision Computing*, **18**(2), 173–187.

Pass, G. and Zabih, R. (1999) Comparing images using joint histograms, *Multimedia Systems*, **7**(3), 234–240.

Puzicha, J., Rubner, Y., Tomasi, C. and Buhmann, J. (1999) Empirical evaluation of dissimilarity measures for color and texture, in *Proceedings the IEEE International Conference on Computer Vision (ICCV99)*, 1165–1173.

Rauber, A. and Merkl, D. (1999) The SOMLib digital library system, in *Proc. of European Conference on Digital Libraries*, 323–342.

Ritter, H. (1991) Asymptotic level density for a class of vector quantization processes, *IEEE Trans. Neural Networks*, **2**, 173–175.

Rubner, Y., Tomasi, C. and Guibas, L. (1998) A metric for distributions with applications to image databases, in *Proceedings of the Sixth International Conference on Computer Vision*, 59–66.

Santini, S. and Jain, R. (1998) Beyond query by example. In *Proceedings of ACM Multimedia 98*, Bristol, UK, ACM Press, 345–350.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A. and Jain, R. (2000) Content-based image retrieval at the end of the early years, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **22**(12), 1349–1380.

Smith, J. and Chang, S. (1996) VisualSEEK: a fully automated content-based image query system, in *Proceedings of ACM Multimedia 96*, Boston, ACM Press, 87–98.

**Received:** Aug. 20<sup>th</sup> 2004

**Accepted in final format:** Nov 20<sup>th</sup> 2004

**About the author:**

Da Deng received his B.Sc. degree from the University of Electronic Science and Technology, Chengdu, PR China in 1989, his M.Sc. and Ph.D. degrees from South China University of Technology (SCUT) in 1992 and 1995 respectively. From 1993 to 1995 he worked as a Research Assistant at the University of Hong Kong. He was a lecturer at SCUT from 1995 to 1999. He has been with the University of Otago, New Zealand since 1999. Dr Deng's research interest includes image analysis, neural networks, and computer networking. He can be reached by email at [ddeng@infoscience.otago.ac.nz](mailto:ddeng@infoscience.otago.ac.nz)