Object Based Image Ranking Using Neural Networks

Gour C. Karmakar[†], Syed M. Rahman[‡], and Laurence S. Dooley[†]

[†] Gippsland School of Computing & Information Technology, Monash University
Gippsland Campus, Churchill VIC 3842, Australia
E-mail: {Gour.Karmakar, Laurence.Dooley}@infotech.monash.edu.au

[‡] Department of Computer and Information Sciences, Minnesota State University
Mankato, MN56001, USA
E-mail: mahbubur.syed@mnsu.edu

Abstract. In this paper an object-based image ranking is performed using both supervised and unsupervised neural networks. The features are extracted based on the moment invariants, the run length, and a composite method. This paper also introduces a likeness parameter¸ namely a similarity measure using the weights of the neural networks. The experimental results show that the performance of image retrieval depends on the method of feature extraction, types of learning, the values of the parameters of the neural networks, and the databases including query set. The best performance is achieved using supervised neural networks for internal query set.

1 Introduction

With the emergence of the Internet and multimedia technology the development of effective content-based image retrieval systems has become a major research area. The Internet, through the World Wide Web, has given users the capability to access vast amounts of information stored on image databases across the network [1]. Object ranking is defined as the listing of objects in decreasing order of similarity to a query object, and is performed by retrieving objects similar to the query reference. The similarity is measured by matching the two objects. There are two types of matching – exact matching and partial matching [2]. Exact matching retrieves the items that are perfectly matched with the user request while partial matching retrieves those items where there are some similarities, but it is not a perfect match. Exact matching is usually utilized in relational databases to perform a query [3], while partial matching, because of the huge variance in pictorial information is used in image retrieval systems. Ranking therefore represents the degree of the resemblance of objects in the database with the query object. A higher order of ranking denotes a higher degree of likeness.

In this paper a model is introduced using neural networks to measure the similarity between the objects in a database and a query object. The ranking is performed using the actual value of the output node of the neural network [1]. Features were extracted from the objects by utilizing three types of feature extraction methods: the moment invariants [4], the run length [5], and the composite [6] methods. Experiments were conducted to compare the discriminating capability of the

features obtained from the composite, the run length and the moment invariants methods. A comparative study was also performed to examine the performance of supervised and unsupervised modes of training for neural networks for image based ranking applications.

The organization of the paper is as follows. A discussion of the proposed models for similarity measurement based on neural networks is given section 2. Section 3 explains the techniques of object ranking using neural networks. The methods for retrieval performance evaluation for object ranking and experimental results are illustrated in sections 4 and 5 respectively. Finally some concluding remarks are presented in section 6.

2   Similarity Measure Using Neural Networks

Various object features are used to train the neural networks. During training, the weights of the networks are adjusted, representing the knowledge about particular object' s features. The actual value of an output unit is determined by applying an activation function to the values of the weights connected to that output unit and this determines the class into which the object is classified. The value of the network's output unit can be used to measure the similarity between the two objects. The similarity measurement is described by the following formula:-

$$Sim(O_o, Q_o) = \left(1 - \frac{|O_o - Q_o|}{Q_o}\right) \times 100\% \tag{1}$$

where O_o is the actual value of network's output unit for the database object;

Q_o is the actual value of network's output unit for the query object;

$Sim(O_o, Q_o)$ is the similarity measure between O_o and Q_o.

This approach is simple, computationally cost effective and can be easily implemented since the neural network has been chosen as the classifier to perform the ranking. It also gives good performance in measuring the similarity but the main disadvantage is that its performance depends on the neural network correctly classifying the object. If the object is misclassified the output value is totally irrelevant.

3   Techniques of Object Ranking Using Neural Networks

Artificial neural networks may be employed for classification of objects. In the first step the database objects are classified using a neural network. Then the class to which the query object belongs is identified and the ranking of the objects within that class with respect to the query object is computed. The detailed steps for ranking by an artificial neural network are shown in Fig. 1. Query object(s) are presented to the system in order to identify similar objects and their ranking. Features are obtained from both the query object(s) and the database objects by the feature extracting methods. Both types of neural networks (supervised and unsupervised) are used for ranking of objects. Neural networks are trained using the features set from the database objects and the classification of the objects contained in the database is performed using the same features set used in training. The feature set computed from
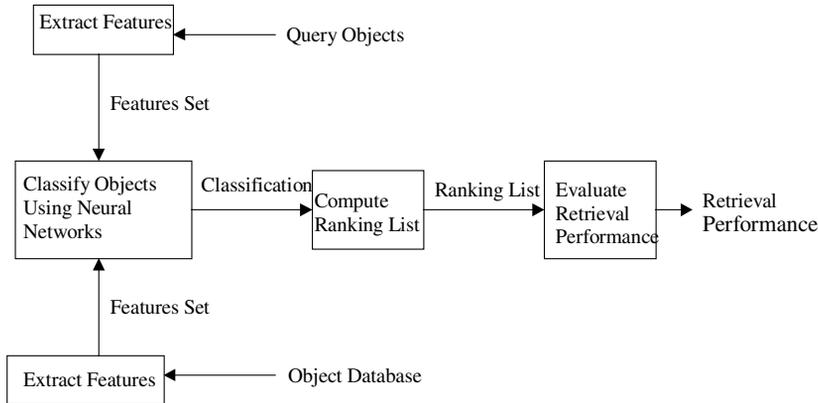
Fig. 1. Block diagram of steps necessary for object ranking using neural networks approach

the query object(s) is applied to the neural networks in order to get the class(es) to which the query object(s) belong. The similarity between the query object(s) and the database objects is computed using the neural network based and city block distance function [7] approaches. The rank order among the objects is derived using the similarity values. The performance of retrieval for ranking purposes is evaluated using the retrieval performance evaluation approaches described in Sect. 4.

4   Methods of Retrieval Performance

Performance evaluation mainly focuses on two aspects, namely effectiveness and efficiency [2]. Effectiveness reflects the capability of the system to satisfy the user whereas efficiency measures the cost and the effort required to solve a specific problem [9]. Cleverdon gave six measurable quantities in order to measure the performance of information retrieval systems [8]. They are the coverage of collection, the average retrieval time (time lag) i.e. the time between the search request and the retrieved objects, the presentation of the output, the endeavor needed for the user to get output to his search request, the recall and the precision of the system. The first four criteria can be easily measured. Recall and precision are most commonly used to measure the effectiveness of a retrieval system. The relevance of the image is determined by the user's satisfaction level based on the significance of the image to a query image. Recall is defined as the ratio between the number of relevant images retrieved and the total number of relevant images in the image database. Precision is defined as the ratio between the number of relevant images retrieved and the total number of images retrieved [10]. The output of the retrieval system can be presented in various ways such as rank position or co-ordination level (the number of common terms between the query and images) [8]. The precision-recall value can be represented by an ordered pair (R, P), if the output of the retrieval system is various ranks as cut-off points. In order to measure the overall performance of the information retrieval system, the average values of recall and precision are computed from all the results for a set of queries. We can compute the average values in two main ways,

average macrorecall and average macroprecision [11][2]. The average macrorecall and macroprecision can be defined by the following equations,

$$\text{Average macrorecall} = \frac{1}{N}\sum_{i=1}^{N}\frac{S_i}{S_i + T_i} \tag{2}$$

$$\text{Average macroprecision} = \frac{1}{N}\sum_{i=1}^{N}\frac{S_i}{S_i + U_i} \tag{3}$$

where   N is the number of queries;

S_i is the number of relevant objects retrieved in ith query;

T_i is the number of relevant objects not retrieved in ith query;

U_i is the number of non-relevant objects retrieved in ith query.

The average values of macro recall-precision pairs obtained from the above pair of equations are used to draw the average curve for recall-precision pairs.

5   Experimental Results

The object database contains around 200 different real objects having variation in shape, scale and orientation. The objects are classified into nine manual classes based on the model of the object. Four query sets were used in the experiments. Each query set consists of nine real objects and each object is chosen from each manual class. A representative sample of query sets (query set 1) after rotation normalization used in object ranking based on neural networks is shown in Fig. 2. The experiments were conducted using both supervised and unsupervised neural networks.

Fig. 2. Query set 1 after rotation normalization

5.1   Ranking Using Supervised Neural Networks

Internal and external queries were used for supervised neural networks. The internal query set (query set 1) is a subset of learning set whereas the external query set is taken from the outside of the learning set. Back propagation network having three slabs in the hidden layer with different activation functions have been applied for supervised leaning. The architecture of the network and its parameters used in training are the following.

Parameters used in the run length and the composite method for internal query set:

  Number of neurons in each hidden slab=30
  Minimum average error=0.00005

Learning epochs taken (run length) =3662
Learning epochs taken (Composite) =152
Learning rate = 0.1
Momentum=0.1
Initial weights=0.3
Activation functions linear [-1,1], Gaussian, Tanh, Gaussian-complement and
logistic are used for input, three hidden and output units respectively.

Parameters used in the moment invariants method for internal query set:

Number of neurons in each hidden slab=8
Minimum average error=0.0016541
Learning epochs taken =307384
All other parameters are the same as the run length and composite methods.

The macro recall-precision curves obtained from the experiments are shown in Fig. 3. The experimental results show that the composite method clearly outperforms both the run length and moment invariants methods. The composite method gave 100% precision for all values of recall, which demonstrates that features obtained from this method, have a good discriminatory power. The main drawback of this method is that it is not fully automatic, as the user needs to add the query set into the object database and assign the class to which the query belongs and then perform the query. The similarity values for this method were computed utilizing the proposed neural networks approach (equation 1) i.e. the actual output values of the neural network.

Another series of experiments were performed using three different external query sets, each of which contained nine different real objects. These query sets were not subsets of the learning set and different numbers of neurons in each hidden slab were used for the run length and the composite methods. A representative samples of the experimental results and parameters used in training are shown in Tables 1 and 2. Minimum average error was taken as 0.0005 and the similarity values were computed using the neural networks approach. The remaining parameters were exactly the same as the parameters used in the run length and composite methods for an internal query set. The experimental results show that the values of recall precision vary with the query set even for the same number of neurons in each hidden slab and other parameters. The selection of query set and number of neurons in each hidden slab have an impact on the precision. If the correlation coefficient of the learning set tends to 1, the precision will be determined only by the correct classification of the query set and it remains constant for all recall values. For this region, the precision values of the composite method are constant for some neurons in Table 2. In the composite method, the maximum precision obtained was 0.67 for data set 2 with 40 neurons in each hidden slab. In the run length method, it was 0.65 for data set 3 with 15 neurons in each hidden slab. The computational time increases with an increase in the number of neurons in each hidden slab. The results also show that the required number of epochs and time during learning for the composite method is much less than the run length method. Thus it may be concluded that the composite method is better than the run length method with respect to precision, learning time and total number of epochs required. If users require the same precision irrespective of recall then this method

may be applied in practical applications but it is very hard to obtain 100 percent precision even for a low level (i.e. 10 percent) of recall.

Fig. 3. Macro recall-precision curve averaged for an internal query set (query set 1) using supervised neural networks

Table 1. The precision values at various levels of recall for query set 1 for the run length method using supervised neural networks

#neurons / hidden slab	Number of epochs	Learning time hh:mm:ss	Precision at various levels of recall (in percentage)									
			10	20	30	40	50	60	70	80	90	100
15	2051	00:28:39	.35	.33	.33	.33	.34	.34	.33	.34	.34	.34
20	1820	00:32:32	.52	.46	.46	.46	.46	.45	.46	.45	.45	.45
25	1674	00:37:08	.52	.54	.54	.54	.55	.55	.55	.54	.54	.54
30	1827	00:47:22	.44	.44	.43	.43	.44	.44	.44	.43	.43	.43
40	1732	00.59:09	.22	.20	.21	.21	.21	.21	.22	.22	.22	.22

Table 2. The precision values at various levels of recall for query set 1 for the composite method using supervised neural networks

#neurons / hidden slab	Number of epochs	Learning time hh:mm:ss	Precision at various levels of recall (in percentage)									
			10	20	30	40	50	60	70	80	90	100
15	66	00:01:00	.35	.35	.35	.35	.35	.35	.35	.35	.34	.34
20	74	00:01:28	.34	.34	.34	.34	.34	.34	.34	.34	.34	.34
25	68	00:01:42	.34	.34	.34	.34	.34	.34	.34	.34	.34	.34
30	75	00:02:15	.28	.28	.24	.24	.24	.24	.24	.24	.24	.24
40	68	00:02:41	.44	.44	.44	.44	.44	.44	.44	.44	.44	.44

5.2   Ranking Using Unsupervised Neural Networks

We have applied unsupervised neural networks, called self-organising maps (SOM), in order to retrieve similar objects and rank them with respect to the user's query. The experiments have been conducted having considered three different numbers of output units i.e. 9, 10 and 15 as the classification results vary with the number of output units of the self-organising map. The values of the parameters used during training of the SOM are the following:

- Initial weight = 0.50, learning rate = 0.50, and neighbourhood radius is taken as the number of output neurons – 1
- The distance function used was Vanilla (Euclidean) and feature selection method is Rotation
- A missing value is regarded as error condition and the total number of epochs is 10000.
- Winning neuron = 1 and all others = 0

Objects are classified by applying both learning and a test set to the SOM. Three different query sets were used for this purpose. The similarity values were calculated using the city block distance function. The average values of macro-precision were computed using equation 3 for 10% to 100% recall levels. Fig. 4 shows the effectiveness of object retrieval in terms of macro recall-precision of the composite, run length and moment invariants methods for 9 output neurons. The same experiments were performed using 10 and 15 output neurons, but this has not been included in this paper.
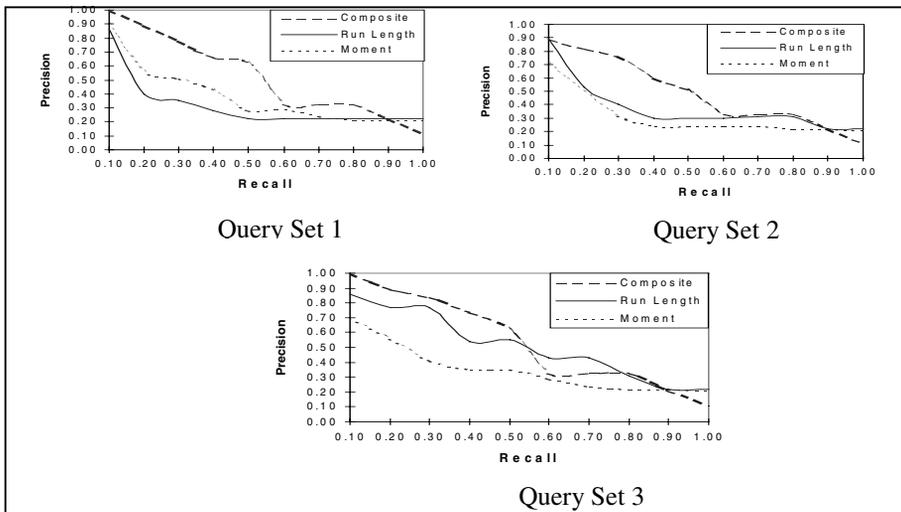

Fig. 4. Average macro recall precision curves for 3 differ

The macro recall-precision curves show that the composite method gave significantly improved precision over the run length and moment invariants methods up to approximately 60% of recall for almost all query sets and for different numbers of output units. The composite method gave poor performance for a higher level of recall, between 80% and 100% of recall compared to the run length and moment invariants methods.   This is due to the change in orientation angle due to the variations of the pixels of the objects of the same class, resulting in changes in the distribution of horizontal chord lengths. Besides, the shapes of the objects in the same class are also different to some extent and manual segmentation adds to this. However, the overall classification performance of the composite method was better than that of both the run length and the moment invariants methods. As a result the composite method only gave poor performance for higher levels of recall, but overall the composite method based on SOM achieved better performance up to 50% of recall when the number of output units is equal to the manual classes (i.e. 9). The run length also outperformed the moment invariants method. We also analyzed the experimental results for each method considering different sets of query objects and the number of output units (9, 10 and 15). The results for 15 output neurons are shown in Fig. 5. The graphs showed the sensitivity of the different feature extraction methods to variations in the query sets and the number of output units used.
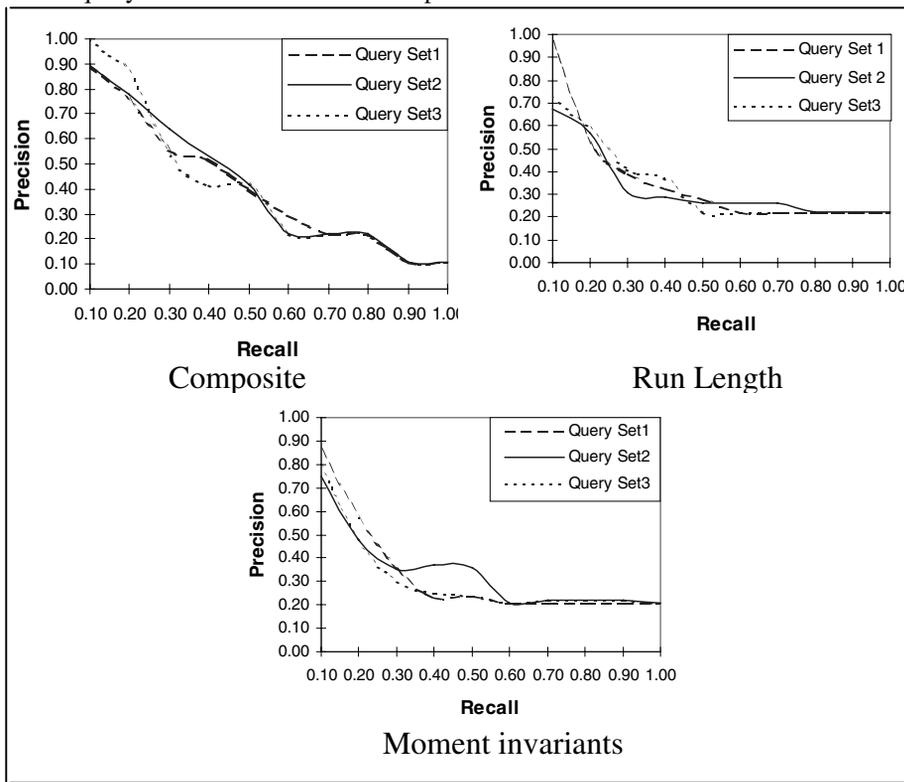

Fig. 5. Average macro recall precision curves for 3 feature extraction methods for 15 output neurons

The recall precision curves show that the composite method can cope better with variation in the query objects than the run length and moment invariants methods. The run length based features are more sensitive to variations in the query set compared to the moment invariants method. So it can be concluded that the composite features have significant discriminating potential in object recognition. Unsupervised neural based object ranking will give promising results for practical applications when users need to retrieve up to 50% of the total relevant objects contained in a database. It is not practically feasible to rank the objects using the traditional computational method without grouping the objects or retrieving some of the objects for a given criterion, for a database containing a large number of objects.

6   Conclusion and Discussions

In this paper we have investigated three feature extraction methods: namely the composite, the run length and the moment invariants methods. Four query sets have been utilized for ranking purposes. Each query set consisted of nine different real objects and was selected from the object database. Ranking was performed using the achieved classification output of the neural networks. We have experimented by applying two types of neural networks, i.e. supervised and unsupervised, to rank the objects matched to a user's query. The similarity values were computed by applying the distance function method (city block distance) and proposed neural networks approach. The neural networks approach gave good similarity values with respect to the query object. Supervised neural networks achieved better performance when the query set is a subset of the learning set but the major drawback is that it needs manual intervention to assign classes to all query objects and add them to the database before performing the query. The precision values (Tables 1 and 2) vary with the number of neurons in each hidden slab and the query set used for supervised neural networks based ranking when the query set is not a subset of the learning set. For some experiments, the classification correlation coefficient of the learning set approached unity and in these cases the precision remains constant irrespective of recall levels. The time and total number of epochs required for training in the composite method was very small compared to the run length method. The experiments on ranking have also been conducted for each feature extraction method using unsupervised neural networks. In this case, the ranking performance was promising and suitable in practice for large databases. The composite method gave a poorer performance for high level of recalls compared with the run length and the moment invariants methods due to errors in rotation normalization, manual segmentation and differences in object shape in a particular manual class. In conclusion, the composite method gives the best performance of the three feature extraction methods and the overall performance of the run length method is better than the moment invariants method.

References

1.   Karmakar, G. C.: Shape And Texture Based Feature Extraction For Object Ranking Using Neural Networks. Masters Thesis, Gippsland School of Computing and Information Technology, Monash University, Australia, September 1998.

2.   Chen, C.: Image Retrieval Using Multiresolution Analysis and Wavelet Transform. Masters Thesis, Department of Computer Science, RMIT University, Australia, March 1997.
3.   Gudivada, V. N., Raghavan, V. V.: Content Based Image Retrieval Systems. IEEE Computer Magazine, pp. 18-22, September 1995.
4.   Ming-Kuei Hu. Visual Pattern Recognition by Moment Invariants. IRE transactions on Information Theory, Vol: IT-8, Feb 1962.
5.   Karmakar, G. C., Rahman, S. M., Bignall, R.J.: Object Ranking Using Run Length Invariant Features. In the Proceedings of the International Symposium on Audio, Video, Image Processing and Intelligent Application (ISAVIIA 98), Baden-Baden, Germany, pp: 52-56, August 1998.
6.   Karmakar, G. C., Rahman, S. M., Bignall, R.J.: Composite Features Extraction and Object Ranking. In the Proceedings of the International Conference on Computational Intelligence for Modeling, Control and Automation (CIMCA'99), Vienna, Austria, pp: 134 - 139, 17-19 February 1999.
7.   Rahman, S. M., Haque, N.: Image Ranking Using Shifted Difference. In Proceedings of the ISCA 12th International Conference on Computers and Their Applications. Tempe, Arizona, USA, March, pp. 110-113, 1997.
8.   Rijsbergen, V. C. J.: Information Retrieval. Butterworths, London, Second Edition, 1997. http://www.dcs.glasgow.ac.uk/Keith/Chapter.7/Ch.7.html.
9.   Salton, G., McGill, M. J.: Introduction to Modern Information Retrieval. McGraw-Hill, New York, 1983.
10.   Sajjanhar, A., Lu, G.: A Grid-Based Shape Indexing and Retrieval Method. The Australian Computer Journal, Vol: 29, No: 4, pp: 131-140, November, 1997.
11.   Salton , G.: Automatic Information Organisation and Retrieval. McGraw-Hill, New York, 1968.