A remote-sensing image-retrieval model based on an ensemble neural networks

Caihong Ma\textsuperscript{a,b}, Fu Chen\textsuperscript{b}, Jin Yang\textsuperscript{b}, Jianbo Liu\textsuperscript{b}, Wei Xia\textsuperscript{b} and Xinpeng Li\textsuperscript{b}

\textsuperscript{a}Sanya Institute of Remote Sensing, Sanya, Hainan Province, China; \textsuperscript{b}Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China

\section*{ABSTRACT}

With the rapid development of remote-sensing technology and the increasing number of Earth observation satellites, the volume of image datasets is growing exponentially. The management of big Earth data is also becoming increasingly complex and difficult, with the result that it can be hard for users to access the imagery that they are interested in quickly, efficiently and intelligently. To address these challenges, this paper proposes a remote-sensing image-retrieval model based on an ensemble neural networks. This model can make full use of existing training data to improve the efficiency and accuracy of the initial retrieval of remote-sensing images and keep model simple. The retrieval of aerial images using the proposed model is compared with the results obtained using ten individual neural networks and two ensemble neural networks and the results show that the proposed approach has a high degree of precision. In addition, the coverage rate and mean precision show a dramatic improvement of more than 40\% compared with existing methods based on normal way. And, the coverage ratio gets 86\% for the top 10 return results.

\section{1. Introduction}

With the development of remote-sensing technology, the volume of image data that is received from satellites has become huge (Liu, Yang, Chen, Dai, & Zhang, 2014; Yasar, Hatipoglu, & Ceylan, 2015). The widespread availability of high spatial resolution remote-sensing images is not only producing an explosion in the volume of acquired data but the amount of detail in the imagery is also increasing by orders of magnitude (Datcu et al. 2005; Wang, Shao, Zhou, & Liu, 2014). This is collectively called big Earth data (Guo, 2017). However, limited by data processing and analysis capacities, the management of big Earth data has not kept up with the rapid increase in the amount of remote-sensing imagery. State-of-the-art systems for accessing remote-sensing images often rely on keywords or tags that relate to geographical coordinates, the data acquisition time or the sensor type (Ma, Dai, Liu, Liu, & Yang, 2014). These keywords or tags may be less relevant than, for example the content of the scene, the structures, patterns or objects that it contains, or the related scattering properties, meaning that users cannot always obtain the images which they are really interested in. Highly efficient solutions based on new technology are,
therefore, considered necessary for easy access to the content of remote-sensing images required by users (Ferecatu & Boujemaa, 2007; Li, Yu, & Yuan, 2016).

Content-based image retrieval (CBIR) has proved to be a major breakthrough in this field. In contrast to the keyword-to-find-image approach, CBIR aims to search images by using visual features that are similar to those in the query image that users submit from big Earth data. It uses a description of automatically extracted visual features, such as colour, texture and shape. After a user submits one or more query images, images in a database are ranked according to their similarity to the query images and the most similar images are returned to the user (Marakakis, Galatsanos, Likas, & Stafyllopatis, 2011). As an efficient way of managing and using the information in an image database from the point of view of comprehension of the image content, the CBIR technique provides a new way of solving the problem of information management in a large remote-sensing image database (Ma et al., 2014; Zhang, 2008). Content-based remote-sensing image retrieval (CBRSIR) has, therefore, attracted the attention of scholars around the world. This will become particularly important in the next decade when the number of acquired remote-sensing images dramatically increases. Feature extraction is the basis of CBIR. In the published CBRSIR literature, several primitive features for characterizing and describing images have been presented for retrieval purposes; these include the fuzzy colour histogram (Han & Ma, 2002), the integrated colour histogram (Hsu, Chua, & Pung, 2003), the Gray Level Co-occurrence Matrix (GLCM) (Ojala, Pietikainen, & Maenpaa, 2000), the fast wavelet (Cheng, 2005) and visual salient point features (Wang et al., 2014). However, most of these studies focus on methods related to different visual features and their effects on CBRSIR (Zhao et al., 2012) and a single feature type cannot always express the image content precisely and perfectly (Wang, Yang, & Li, 2013); it is also hard to obtain satisfactory retrieval results using a single feature. Therefore, in this paper, the multi-feature (MF) integrated retrieval model is proposed, in which three categories of colour features and four categories of texture features are applied to the retrieval of remote-sensing images in order to improve the image retrieval.

The ability of the low-level features of a remote-sensing image to describe the semantic content is also very limited. This is a well-known semantic gap problem that occurs between the low-level features and the high-level semantic content and leads to the intrinsic difficulty in capturing the human perception of image similarity for CBIR. This gap seriously limits the success of CBIR (Liu et al., 2007). To bridge this gap, since the 1990s, most studies of CBIR have also included a post-interactive process named relevance feedback (RF) and this approach has achieved considerable success (Ma et al., 2014; Xiang & Huang, 2003). During rounds of RF, users are required to label the retrieved images as relevant or irrelevant to the query image. Then, the retrieval system takes into account the users’ feedback to update the ranking criterion (Mountrakis, Im, & Ogole, 2011). Many studies show that the retrieval results are greatly improved after two or three rounds of feedback. However, RF is an after-the-fact feedback mechanism and need a lot of manual intervention; it is also unable to change the initial retrieval results. And, some researchers introduced deep learning methods on image retrieval (Li, Zhang, & Huang et al., 2017; Wan, Wang, & Hoi et al., 2014; Zhou, Newsam, Li, & Shao, 2017). Those experiments results show that they all improve the accuracy of remote-sensing image retrieval on the same training sample.
set. However, they must propose new models to keep the volume of training data. It increases the complexity of complex deep learning models and decline its universality on other datasets. Therefore, how to improve the efficiency and accuracy of the initial retrieval of remote-sensing images without any human intervention and keep simple is becoming an important question. Predictive models built from “experience” – which, in practice, means data acquired from actual cases – provide a feasible solution (Dreiseitl & Ohno-Machado, 2002). The neural network, a black-box predictive model, has been utilized to improve image classification (Elalami, 2014). It can be applied to CBIR when images in the archived database have different labels. One neural network cannot always adapt to all kinds of classification, especially in cases where the number of categories is large. Therefore, how to construct a group of base learners to improve the efficiency and accuracy of initial remote-sensing initial retrieval is still an important question.

To overcome the problems discussed above, a remote-sensing image retrieval model based on an ensemble neural networks (ENNs) is proposed in this paper. According to the new model, it includes two main procedures: training neural networks with different subfeatures and the content-based remote-sensing image retrieval. In the training neural networks processing, different neural networks were trained by selecting different features with same neural network framework and training data. And, query image can be assigned its posterior probabilities, calculated by ENNs, for different classes in image retrieval part. Then, the posterior probabilities can then change the traditional similarity. Therefore, images in which the similarity between the classes is higher will have larger posterior probabilities and, as a result, a much higher similarity. The new model makes two general improvements to the content-based remote-sensing image retrieval model. Firstly, the new model makes full use of the existing training data to construct a neural network and then improve the efficiency and accuracy of the initial image retrieval. Secondly, an ENNs model for different feature subsets is used to improve the stability of the initial retrieval results.

This paper is organized as follows. Section 2 describes related algorithms, including feature extraction and the basic artificial neural network (ANN) concept. Section 3 describes in detail the remote-sensing image-retrieval model based on the ENNs. Section 4 presents the experimental results that were obtained using the aerial images database dataset. The experimental results are also compared with the results of using traditional approaches. Conclusions are finally drawn in Section 5 and are presented along with recommendations for future research.

2. Review of related algorithms

2.1. Feature extraction in remote-sensing images

As colour is insensitive to image rotation and translation, as well as to image size and direction, it is considered to be the most expressive type of visual feature and has been extensively studied. Texture is resistant to noise and is invariant under rotation; textural patterns are also scale invariant. Texture is also regarded as an important visual feature in relation to the innate surface properties of a ground object and their relationship to the surrounding environment. In this study, therefore, three colour features and four texture features were extracted from remote-sensing images.
In colour feature extraction, the colour of a pixel is usually given as three values corresponding to $R$ (red), $G$ (green) and $B$ (blue). The colour histogram, which indicates the frequency of occurrence of the different colours in the image, is the most common description of the colour. Types of colour histogram include the integrated colour histogram (Hsu et al., 2003) and the fuzzy colour histogram (Han & Ma, 2002). The colour correlogram presented by Huang, Kumar, and Mitra (1997) expresses clearly how the spatial correlation of the colours changes with distance. As it yields a better retrieval accuracy than the colour histogram, it is adopted in many CBIR systems. However, $R$, $G$ and $B$ have a strong intensity factor and heavily relativity to each other, and so, the HSV–HIST histogram description was proposed (Liu & Zhang, 1998). In addition, the colour moments description is a simple global algorithm that has been used to describe the global visual features.

In texture feature extraction, there are some conventional statistical texture features such as the GLCM, the Markov random field model and the edge histogram descriptor (Wang et al., 2013). As the GLCM (Ojala et al., 2000) can measure properties such as entropy, correlation and contrast well, this description is one of the most well-known and widely used texture visual features. In order to take the space of all possible intensity patterns in a neighbourhood into consideration, the Texture Spectrum (Topi, Matti, & Timo, 2000) feature is also used. Moreover, local binary patterns (Ojala et al., 2000) an improved Texture Spectrum, has been proposed to improve the ability of anti-direction and anti-overturn. The improved Texture Spectrum description also quantizes the traditional 256-dimensional texture spectrum description into 51 dimensions. The fast wavelet (Cheng, 2005), which is non-separable and oriented, also improves the characterization of diagonally oriented textures. In addition, in texture feature extraction, in-moments (Ma et al., 2014) is also commonly used as it is insensitive to image size and direction.

### 2.2. Artificial neural network

It is known that an ANN can automatically explore, create and derive new information by learning without any help. An ANN incorporates large-scale parallel computation, distributed processing, self-organization and self-learning and so is particularly useful in applications where the complexity of the data or task makes the design of such a function by hand impracticable. ANNs have been widely used in many fields such as speech analysis, image recognition, digital watermarking and computer vision and have achieved many outstanding results (Yasar et al., 2015). Recently, due to their rapid development, ANNs have become a powerful tool for pattern recognition. Although there are different types of ANN, feed-forward back propagation (BP) ANNs are the most widely used type as they have the advantage of being able to deal effectively with the exclusive-or problem and, more generally, the problem of quickly training multi-layer neural networks. An BP ANN is an artificial intelligence algorithm consisting of an input layer, a hidden layer and an output layer, as shown in Figure 1.

Assuming $n$ denotes the number of input neurons, it also represents the $n$-dimensional feature vector in this paper. $m$ is the number of output neurons and also denotes the number of target classes. $X = (x_1, x_2, \cdots, x_n)$ represents one input and $Y = (y_1, y_2, \cdots, y_m)$ the corresponding output. Therefore, $f : X \rightarrow Y$ is the BP ANN learning function. A BP ANN usually contains three layers as mentioned before. The input layer and the output layer each consist of
one layer, but the hidden layer can be set to have one or more layers according to the needs of the network structure – there can also be no hidden layer. In the BP ANN model, the input layer, hidden layer and output layer are interconnected by means of weights. Forward stimulation is used to reset the weight coefficients on the front neural base to the minimal error value between the target and the output that is produced.

3. A remote-sensing image-retrieval model based on an ensemble neural network

In this study, we propose a remote-sensing image-retrieval model based on an ENN. The proposed approach makes full use of the existing training data to improve the efficiency and accuracy of the initial remote-sensing image-retrieval results. At the same time, an ENN model for different feature subsets is proposed to improve the stability of the retrieval results. The main procedures that comprise the remote-sensing image-retrieval model are shown in Figure 2. As shown in the figure, the new model contains two main parts: the training of the ANN and the content-based remote-sensing image retrieval.

The ANN training consists of the following steps:
Step 1. Select the training data.
Step 2. Extract the feature vector of the training data and scale the value to the range $[0, 1]$.
Step 3. Select sub-features for different ANNs.
Step 4. Construct the ANN structure. This includes setting the number of hidden layers in the neural network as well as the number of neurons per hidden layer.
Step 5. Train the ANN. In this study, the training data selected in step 1 were used to train the ANN structure that was constructed in step 3. The BP learning mechanism was used for this training.
Step 6. Form the ANN set. One ANN model will be trained with a kind of sub-features.

In this paper, sub-features were comprised category of features or combined some categories of features.

The main steps in the content-based remote-sensing image retrieval are as follows:
Step 1. Input the target image.
Step 2. Extract the feature vector of the target image and scale the value to the range $[0, 1]$. 

Figure 1. The Schematic diagram of ANN.
Step 3. Calculate the posterior probabilities for the feature vector of the query image for different classes. In this study, the posterior probability is the sum posterior probabilities calculating by different ANN models in ANN set. This can be calculated using Equation (1):

$$P_i = \sum_{j=1}^{M} p_{ij}$$  

where, $M$ is the number of sub-ANN models. In this paper, the value of $M$ is 10. $p_{ij}$ represents the posterior probability for the $i$th class for the $j$th sub-ANN model. It is calculated from the classification outputs of the ANN. $P_i$ represents the total posterior probability for the $i$th class.

Step 4. Calculate the distance between the feature vector of the query image and the feature database. The Euclidean distance was used for this.
Step 5. Calculate the feature similarity between the feature vector of the query image and the feature database. The similarity is calculated using Equation (2):

$$D_k = \frac{d_k}{P_i}$$  

where, $d_k$ represents the Euclidean distance between the feature vector of the query image and the $k$th image in the feature database. The $k$th image belongs to the $i$th class. $P_i$, calculated in step 3, represents the total posterior probability for the $i$th class. $D_k$ is the similarity between the feature vector of the query image and the $k$th image.

Step 6. Sort the feature similarity between the feature vector of the query image and the feature database according to $D_k$.

Step 7. Return the top $N$ similar images. In our experiment, the value of $N$ was 21.

4. Experiments

4.1. Remote-sensing image database and feature database

4.1.1. Remote-sensing image database

In order to assess the effectiveness of the proposed remote-sensing image-retrieval model, we carried out experiments using a database of aerial images that consisted of characterizing 21 land-use classes (Yang & Newsam, 2010). Large images selected from a database of aerial ortho-imagery were downloaded from the USA Geological Survey (USGS) National Map of the following US regions: Birmingham, Boston, Buffalo, Columbus, Dallas, Harrisburg, Houston, Jacksonville, Las Vegas, Los Angeles, Miami, Napa, New York, Reno, San Diego, Santa Barbara, Seattle, Tampa, Tucson and Ventura. Each land-use class contained 100 images measuring $256 \times 256$ pixels with a pixel resolution of 30 cm. Each image belonged to one of the following 21 land-use classes: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbour, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks and tennis courts. Figure 3 shows three samples for each of the 21 classes.

These images were divided into two sets: one for training the neural network and the other for testing. The training data set consisted of 1680 images (80 images for each class) and the testing data set consisted of 420 images (the remaining 20 images for each class).

4.1.2. Feature database

As mentioned in Section 2.1, in this study, three colour feature categories and four texture features categories were used in the retrieval of remote-sensing images. For the colour feature extraction, the colour correlogram (Huang et al., 1997), colour moments and the HSV–HIST histogram (Liu & Zhang, 1998) were used. For the texture feature extraction, the fast wavelet (Cheng, 2005), in-moments, the GLCM and the Texture Spectrum (Ojala et al., 2000; Topi et al., 2000) were adopted.

As shown in Table 1, a 423-dimensional feature vector consisting of the seven features described above was used to represent the content of the aerial imagery. In order to make the calculation easier and to prevent feature values in higher numerical ranges from dominating those in smaller numerical ranges, the seven feature description categories were normalized to the range [0, 1].
4.2. Performance measurement indexes

Evaluation of the retrieval performance is a crucial part of content-based remote-sensing image retrieval. Many different methods for measuring the performance of a system have been created and used by researchers. In this study, we used the most common evaluation methods, namely recall (or sensitivity) and precision (or specificity).

4.2.1. Coverage ratio

Traditionally, the recall, precision and recall-precision break-even point are commonly applied to assess the effectiveness of retrieval models. However, Schapire, Singer, and Singhal (1998) give very reasonable arguments as the conventional evaluation metrics are not very informative for users of a CBIR system. In particular, the recall index cannot
be calculated by the user until all relevant images have been seen by the user, which is not possible except by means of an exhaustive search. Therefore, the user cannot know how well the image search is going (Ma et al., 2014). In this study, then, the coverage ratio was applied to the remote-sensing image retrieval and used as the performance metric. It can be calculated by using Equation (3):

\[
\text{Coverage ratio} = \begin{cases} 
\frac{n_R}{10} & (10i \leq R), \\
\frac{n_R}{R} & (10i > R) 
\end{cases}
\] (3)

Here, \( R \) is the total number of relevant images in the image database and \( n_R \) is the number of relevant images returned for the top \( 10i \) images. When the value of \( 10i < R \), the coverage ratio is the same as the precision; if the value of \( 10i > R \), the coverage ratio is the same as the recall. In this study, \( i \) was set to \{1,2,3,4,5,8\}.

4.2.2. Mean average precision

Precision and recall—precision are single-value metrics that are based on the whole set of images returned by the retrieval system. For systems that return sequences of ranked images, it is desirable to also consider the order in which the returned images are presented. The average precision index emphasizes relevant images by ranking them higher. This index is calculated as the average of the precisions calculated for each of the relevant images in the ranked sequence. The mean average precision for a set of queries is the mean of the average precision scores for each query. This is calculated as:

\[
\text{Map} = \frac{1}{N} \sum_{r=1}^{N} \frac{\rho_s}{\rho_r}
\] (4)

where \( r \) is the rank, \( N_r \) denotes the number of relevant images, \( N_s \) represents the number of real relevant images in the relevant images that are returned, \( \rho_s \) is the rank of the relevant images returned and \( \rho_r \) is the rank of the real relevant images returned.

4.3. Experimental results

Our model was implemented within the Matlab2015a environment. The empirical evaluation was performed on a PC with Dell3G memory and a Win7 operating system. In order to ensure the fairness of comparison, the number of hidden layers was all set as 50. And the number of neurons in first hidden layer was all set as 423. The target number was set as 21.

4.3.1. Statistical results

The results obtained for different feature categories using the remote-sensing image-retrieval model based-on the ENNs were compared with those obtained using 12 other neural networks. In this paper, the ANN set contained 10 ANN models, including ANN models named: colour correlogram, colour moments, HSV–HIST, Fast wavelet, In-moments, GLCM, Texture Spectrum Multi-Colour, Multi-Texture and Multi-all. The colour correlogram, colour moments, HSV–HIST, Fast wavelet, In-moments, GLCM and Texture Spectrum all represent neural networks trained individually using the features described above. Multi-Colour, Multi-Texture and Multi-all are neural networks that combine three types of colour features and four types of
texture feature, a total of seven features. In addition, there were three ensemble models, including: Ensemble Colour, Ensemble Texture and Ensemble All. Ensemble Colour model combined colour correlogram, colour moments and HSV–HIST three ANN models. And Ensemble Texture model combined Fast wavelet, In-moments, GLCM and Texture Spectrum four ANN models. Ensemble All combined all ten ANN models mentioned before.

Tables 2 and 3 show the coverage ratio and mean average precision, respectively, obtained when \( i \) was set to \( \{1,2,3,4,5,8\} \) for 20 trials per category. This resulted in a total of 420 trials using the aerial image database. Tables 2 and 3 show that, for Multi-Colour and Multi-Texture, the coverage ratio and mean average precision were also higher than for the neural networks trained by a single neural network using three kinds of colour feature and four kinds of texture feature. The coverage ratio and mean average precision values for individual neural networks trained by Multi-all was better than for those trained by Multi-Colour and Multi-Texture. This means that training networks using higher-dimensional features produces better results. In addition, the results obtained using Ensemble Colour and Ensemble Texture were both better than for those for individual networks trained by Multi-Colour and Multi-Texture. Overall, Ensemble All produced the highest coverage ratios and mean average precisions.

To validate the results of the proposed method based on an ENN, the results obtained were then compared with the conventional method based on MF. As mentioned earlier, it should be noted that ENN used Ensemble All for the retrieval. Tables 4 and 5 show the coverage ratio and mean average precision, respectively, when \( i \) was set to \( \{1,2\} \) during the 20 trials per category

### Table 2. Comparison of different methods coverage rate.

| Method               | \( i = 1(\%) \)       | \( i = 2(\%) \)       | \( i = 3(\%) \)       | \( i = 4(\%) \)       | \( i = 5(\%) \)       | \( i = 8(\%) \)       |
|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Colour correlogram   | 0.44 ± 0.26           | 0.42 ± 0.26           | 0.41 ± 0.26           | 0.41 ± 0.26           | 0.41 ± 0.26           | 0.40 ± 0.26           |
| Colour moments       | 0.39 ± 0.25           | 0.38 ± 0.25           | 0.38 ± 0.24           | 0.37 ± 0.24           | 0.37 ± 0.24           | 0.36 ± 0.24           |
| HSV–HIST             | 0.39 ± 0.17           | 0.37 ± 0.17           | 0.37 ± 0.17           | 0.37 ± 0.17           | 0.37 ± 0.17           | 0.36 ± 0.17           |
| Fast wavelet         | 0.44 ± 0.23           | 0.42 ± 0.23           | 0.41 ± 0.23           | 0.41 ± 0.23           | 0.41 ± 0.23           | 0.40 ± 0.23           |
| In-moments           | 0.38 ± 0.19           | 0.30 ± 0.18           | 0.26 ± 0.17           | 0.23 ± 0.17           | 0.21 ± 0.16           | 0.18 ± 0.15           |
| GLCM                 | 0.56 ± 0.21           | 0.54 ± 0.21           | 0.53 ± 0.21           | 0.52 ± 0.21           | 0.52 ± 0.21           | 0.51 ± 0.22           |
| Texture spectrum     | 0.69 ± 0.16           | 0.67 ± 0.16           | 0.66 ± 0.17           | 0.66 ± 0.17           | 0.66 ± 0.17           | 0.66 ± 0.17           |
| Multi-Colour         | 0.53 ± 0.18           | 0.51 ± 0.18           | 0.51 ± 0.18           | 0.50 ± 0.18           | 0.50 ± 0.19           | 0.50 ± 0.18           |
| Multi-Texture        | 0.71 ± 0.15           | 0.70 ± 0.19           | 0.70 ± 0.19           | 0.69 ± 0.19           | 0.69 ± 0.19           | 0.68 ± 0.19           |
| Multi-all            | 0.72 ± 0.15           | 0.72 ± 0.15           | 0.72 ± 0.16           | 0.72 ± 0.16           | 0.72 ± 0.16           | 0.71 ± 0.16           |
| Ensemble Colour      | 0.59 ± 0.22           | 0.58 ± 0.22           | 0.58 ± 0.22           | 0.57 ± 0.22           | 0.57 ± 0.22           | 0.56 ± 0.23           |
| Ensemble Texture     | 0.79 ± 0.15           | 0.78 ± 0.15           | 0.77 ± 0.16           | 0.77 ± 0.16           | 0.77 ± 0.16           | 0.77 ± 0.16           |
| Ensemble All         | 0.82 ± 0.13           | 0.81 ± 0.14           | 0.81 ± 0.14           | 0.81 ± 0.14           | 0.81 ± 0.14           | 0.81 ± 0.14           |

### Table 3. Comparison of different methods mean average precision value.

| Method               | \( i = 1(\%) \)       | \( i = 2(\%) \)       | \( i = 3(\%) \)       | \( i = 4(\%) \)       | \( i = 5(\%) \)       | \( i = 8(\%) \)       |
|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Colour correlogram   | 0.48 ± 0.25           | 0.47 ± 0.25           | 0.47 ± 0.25           | 0.46 ± 0.25           | 0.46 ± 0.25           | 0.45 ± 0.25           |
| Colour moments       | 0.44 ± 0.24           | 0.43 ± 0.25           | 0.43 ± 0.25           | 0.42 ± 0.25           | 0.42 ± 0.25           | 0.41 ± 0.25           |
| HSV–HIST             | 0.43 ± 0.17           | 0.42 ± 0.17           | 0.42 ± 0.17           | 0.41 ± 0.16           | 0.41 ± 0.17           | 0.40 ± 0.17           |
| Fast wavelet         | 0.49 ± 0.23           | 0.48 ± 0.23           | 0.47 ± 0.23           | 0.47 ± 0.23           | 0.46 ± 0.23           | 0.45 ± 0.23           |
| In-moments           | 0.58 ± 0.19           | 0.55 ± 0.18           | 0.52 ± 0.18           | 0.50 ± 0.18           | 0.47 ± 0.18           | 0.42 ± 0.17           |
| GLCM                 | 0.62 ± 0.20           | 0.61 ± 0.20           | 0.60 ± 0.20           | 0.59 ± 0.20           | 0.59 ± 0.20           | 0.58 ± 0.21           |
| Texture spectrum     | 0.73 ± 0.14           | 0.72 ± 0.15           | 0.72 ± 0.15           | 0.71 ± 0.15           | 0.71 ± 0.15           | 0.70 ± 0.16           |
| Multi-Colour         | 0.57 ± 0.17           | 0.56 ± 0.18           | 0.55 ± 0.18           | 0.54 ± 0.18           | 0.54 ± 0.18           | 0.53 ± 0.18           |
| Multi-Texture        | 0.74 ± 0.16           | 0.73 ± 0.16           | 0.73 ± 0.17           | 0.72 ± 0.17           | 0.72 ± 0.17           | 0.71 ± 0.18           |
| Multi-all            | 0.75 ± 0.15           | 0.74 ± 0.15           | 0.73 ± 0.15           | 0.73 ± 0.15           | 0.73 ± 0.15           | 0.73 ± 0.15           |
| Ensemble Colour      | 0.65 ± 0.20           | 0.63 ± 0.21           | 0.62 ± 0.21           | 0.61 ± 0.21           | 0.60 ± 0.21           | 0.60 ± 0.21           |
| Ensemble Texture     | 0.83 ± 0.13           | 0.82 ± 0.13           | 0.81 ± 0.14           | 0.81 ± 0.14           | 0.80 ± 0.14           | 0.80 ± 0.14           |
| Ensemble All         | 0.86 ± 0.13           | 0.85 ± 0.13           | 0.84 ± 0.13           | 0.84 ± 0.13           | 0.83 ± 0.13           | 0.83 ± 0.13           |
that were carried out using the aerial image database. According to Tables 4 and 5, the retrieval results obtained using the ENN had higher coverage ratios and mean average precisions than those obtained using MF for all 21 land-use classes in the aerial image database. Also, on average, the coverage rates and mean average precisions obtained using ENN were more than 40 per cent higher than those obtained using MF. For agricultural and airplane land-use classes, the ENN improved the coverage rate by 52 per cent coverage compared with MF. For the intersection class, the mean average precision improved by 51 per cent.

This section summarizes the results obtained for all 21 land-use classes. Figure 4 shows (a) the average precision and (b) recall for each class for the top 20 retrieved results. The results were obtained by selecting 20 query images from each category and then averaging the results. As shown in Figure 4, the performances of the ENN and MF vary according to the category but the proposed method produces a higher average precision and recall for all of the categories. In more detail, the average precision obtained by the proposed ENN was higher than 60% for all classes except (19) sparse residential. Most of the average precisions obtained by MF were lower than 60% for the “difficult” categories, except for (1) agricultural and (11) golf course (11), for which the values were 60.08 and 59.5 per cent respectively.

Figure 5 shows the precision–recall graphs for the proposed ENN and also the MF method that were obtained using 20 trials per category (420 in total) and the aerial

**Table 4.** Retrieval coverage rate comparison between normal way based on multi-features (MFs) and the proposed way based on ensemble neural networks (ENNs) using aerial image database.

| Class          | MF  | ENN |
|----------------|-----|-----|
|                | $i = 1$ | $i = 2$ | $i = 1$ | $i = 2$ |
| Agricultural   | 0.43 | 0.33 | 0.95 | 0.95 |
| Airplane       | 0.41 | 0.31 | 0.80 | 0.80 |
| Baseball diamond | 0.29 | 0.22 | 0.77 | 0.74 |
| Beach          | 0.62 | 0.45 | 0.97 | 0.97 |
| Buildings      | 0.23 | 0.19 | 0.66 | 0.65 |
| Chaparral      | 0.61 | 0.45 | 0.97 | 0.97 |
| Dense residential | 0.41 | 0.30 | 0.82 | 0.80 |
| Forest         | 0.57 | 0.46 | 0.97 | 0.97 |
| Freeway        | 0.22 | 0.16 | 0.93 | 0.92 |
| Golf course    | 0.44 | 0.30 | 0.96 | 0.95 |
| Harbour        | 0.73 | 0.59 | 0.96 | 0.96 |

**Table 5.** Retrieval mean average precision comparison between normal way based on multi-features (MFs) and the proposed way based on ensemble neural networks (ENNs) using aerial image database.

| Class          | MF  | ENN |
|----------------|-----|-----|
|                | $i = 1$ | $i = 2$ | $i = 1$ | $i = 2$ |
| Agricultural   | 0.66 | 0.60 | 0.95 | 0.95 |
| Airplane       | 0.69 | 0.64 | 0.85 | 0.85 |
| Baseball diamond | 0.49 | 0.44 | 0.84 | 0.83 |
| Beach          | 0.85 | 0.78 | 1.00 | 0.99 |
| Buildings      | 0.47 | 0.38 | 0.69 | 0.68 |
| Chaparral      | 0.84 | 0.80 | 0.99 | 0.99 |
| Dense residential | 0.68 | 0.59 | 0.84 | 0.83 |
| Forest         | 0.81 | 0.72 | 0.99 | 0.98 |
| Freeway        | 0.60 | 0.51 | 0.98 | 0.97 |
| Golf course    | 0.82 | 0.72 | 0.99 | 0.97 |
| Harbour        | 0.94 | 0.90 | 1.00 | 0.98 |
Figure 4. Comparison of average precision (a) and recall (b) (on 20 trials per category, resulting to 420 trials in total) obtained for different categories obtained by the proposed way based on ensemble neural networks and normal way based on multi-features when the top 20 images are retrieved: (1-agricultural, 2-airplane, 3-baseball, 4-diamond, 5-beach, 6-buildings, 7-chaparral, 8-dense residential, 9-forest, 10-freeway, 11-golf course, 12-harbor, 13-intersection, 14-medium-density residential, 15-mobile home park, 16-overpass, 17-parking lot, 18-river, runway, 19-sparse residential, 20-storage tanks and 21-tennis courts).
image database. From Figure 5, it can again be observed that the ENN method outperforms the MF technique in terms of the precision-to-recall ratio. Firstly, all of average precisions based on ENN were higher than using MF, with 40%. Then, the

**Figure 5.** Aerial images-retrieval average precision versus recall graphs obtained by the proposed way based on ensemble neural networks and normal way based on multi-features.

**Figure 6.** Query by example: looking for “harbour areas” in a database of aerial images dataset. (a) Query image. (b) Images retrieved by normal way based on multi-features in the case of the top 21 similar images. In this case, the normal way based on multi-feature can be is to provide satisfactory answers.
average precision using MF declined sharply when the average recall increased. While, the average precision based on ENN kept stability.

Based on the above performance comparisons, it can be concluded that the ENN produces better retrieval results than those obtained using one neural network training. The retrieval results obtained using the ENN also show a dramatic improvement compared with the usual method (MF).

4.3.2. Search examples

To illustrate the effectiveness of our approach for the querying of remote-sensing images using a ENN, we provide here some screenshots obtained from our CBRSIR system. Figure 6 shows a typical sample of a query image “harbour areas” (Figure 6(a)) together with the corresponding images retrieved by the MF-based method. The retrieval order for the images is

![Figure 6](image)

**Figure 6.** Query by example: looking for “harbour areas” in a database of aerial images dataset. (a) Query image. (b) Images retrieved by normal way based on multi-features. (c) Images retrieved by the proposed way based on ensemble neural networks in the case of the top 21 similar images.

![Figure 7](image)

**Figure 7.** Query by example: looking for “airplane area” in a database of aerial images dataset. (a) Query image. (b) Images retrieved by normal way based on multi-features. (c) Images retrieved by the proposed way based on ensemble neural networks in the case of the top 21 similar images.
shown in Figure 6(b). It can be seen that most of the images retrieved by the MF-based method look promising, except for the 10th, 19th and 21st images.

As shown in Tables 4 and 5, the MF methods performed better for the harbour land-use class than for other classes. In most cases, the retrieval results obtained by the MF method are not satisfactory and have lower coverage ratios and mean average precisions. Figures 7 and 8 show two other typical queries by visual example (QBE): searches for “airplane area” and “tennis court area” (Figures 7(a) and 8(a), respectively). The order of the query results is the same as that shown in Figure 5. However, for the MF method, the results are not satisfactory as the results of both queries produce the same 8 relevant images in the top 21 images (Figures 7(b) and 8(b)). In contrast, the results obtained using the proposed method all have a precision of 100 per cent, as shown in Figures 7(c) and 8(c).
5. Conclusion

In this paper, we argued that the proposed ENNs can be successfully applied to remote-sensing image retrieval. The retrieval system takes the features of colour and texture into consideration. And the experimental results demonstrate that the proposed ENN obtains better retrieval results than those obtained using one neural network. Also, the results obtained by the ENN show dramatic improvements of more than 40 per cent in the coverage rates and mean average precisions compared with the commonly used MF-based method. And, the coverage ratio gets 86 per cent for the top 10 return results. Overall, the experimental results show that the new simple model can make full use of existing training data to improve the efficiency and accuracy of initial remote-sensing image retrieval.

Data availability statement

The data referred to in this paper is not publicly available at the current time.

Acknowledgements

We would thank to the Hainan Provincial Natural Science Foundation of China (618QN303) and the National Natural Science Funds of China (41501116).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Funds of China [41501116].

References

Cheng, Q. M., Yang, C. J., & Shao, Z. F. (2005). Progressive texture image retrieval based on M-band wavelet features. Geomatics and Information Science of Wuhan University, 30(6), 521–524.
Datcu, M., & Seidel, K. (2005). Human-centered concepts for exploration and understanding of earth observation images. IEEE Transactions on Geoscience & Remote Sensing, 43(3), 601–609.
Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. Journal of Biomedical Informatics, 35(5–6), 352.
Elalami, M. E. (2014). A new matching strategy for content based image retrieval system. Applied Soft Computing, 14(1), 407–418.
Ferecatu, M., & Boujema, N. (2007). Interactive remote-sensing image retrieval using active relevance feedback. IEEE Transactions on Geoscience & Remote Sensing, 45(4), 818–826.
Guo, H. (2017). Big Earth data: A new frontier in Earth and information sciences. Big Earth Data, 1 (1–2), 4–20.
Han, J., & Ma, K. K. (2002). Fuzzy color histogram and its use in color image retrieval. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 11(8), 944.
Hsu, W., Chua, S. T., & Pung, H. H. (2003, November). An integrated color-spatial approach to content-base image retrieval. ACM international conference on multimedia ’95 (pp. 305–313). San Francisco, CA, USA.
Huang, J., Kumar, S. R., & Mitra, M. (1997). Combining supervised learning with color correlograms for content-based image retrieval. ACM international conference on multimedia (pp. 325–334). ACM New York, NY, USA.

Li, S., Yu, H., & Yuan, L. (2016). A Novel approach to remote sensing image retrieval with multi-feature vp-tree indexing and online feature selection. IEEE second international conference on multimedia big data (pp. 133–136). IEEE. Taipei, China. doi:10.1109/BigMM.2016.11

Li, Y., Zhang, Y., Huang, X., Zhu, H., & Ma, J. (2017). Large-scale remote sensing image retrieval by deep hashing neural networks. IEEE Transactions on Geoscience & Remote Sensing, PP(99), 1–16.

Liu, J., Yang, J., Chen, F., Dai, Q., & Zhang, J. (2014). Location-based instant satellite image service: Concept and system design. International Journal of Digital Earth, 8(4), 91–101.

Liu, R., Wang, Y., Baba, T., Uehara, Y., Masumoto, D., & Nagata, S. (2007). Svm-based active feedback in image retrieval using clustering and unlabeled data. Pattern Recognition, 41(8), 2645–2655.

Liu, Z. W., & Zhang, Y. J. (1998). Color image retrieval using local accumulative histogram. Journal of Image and Graphics, 3(7), 533–587.

Ma, C., Dai, Q., Liu, J., Liu, S., & Yang, J. (2014). An improved svm model for relevance feedback in remote sensing image retrieval. International Journal of Digital Earth, 7(9), 725–745.

Marakakis, A., Galatsanos, N., Likas, A., & Stafylopatis, A. (2011). Combining gaussian mixture models and support vector machines for relevance feedback in content based image retrieval. IET Image Processing, 5(6), 31–540.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry & Remote Sensing, 66(3), 247–259.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2000). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis & Machine Intelligence, 24(7), 971–987.

Schapire, R. E., Singer, Y., & Singhal, A. (1998). Boosting and Rocchio applied to text filtering. International ACM SIGIR conference on research and development in information retrieval (pp. 215–223). ACM New York, NY, USA. doi:10.1145/290941.290996

Topi, M., Matti, P., & Timo, O. (2000). Texture classification by multi-predicate local binary pattern operators. International conference on pattern recognition, 2000. Proceedings (Vol. 3, pp. 939–942). IEEE. Barcelona, Spain. doi:10.1109/ICPR.2000.903699

Wan, J., Wang, D., Hoi, C. H., Wu, P., Zhu, J., Zhang, Y., & Li, J. (2014). Deep learning for content-based image retrieval: A Comprehensive study. ACM international conference on multimedia.

Wang, X., Shao, Z., Zhou, X., & Liu, J. (2014). A novel remote sensing image retrieval method based on visual salient point features. Sensor Review, 34(4), 349–359.

Wang, X. Y., Yang, H. Y., & Li, D. M. (2013). A new content-based image retrieval technique using color and texture information. Computers & Electrical Engineering, 39(3), 746–761.

Xiang, S. Z., & Huang, T. S. (2003). Relevance feedback in image retrieval: A comprehensive review. Multimedia Systems, 8(6), 536–544.

Yang, Y., & Newsam, S. (2010, November). Bag-of-visual-words and spatial extensions for land-use classification. ACM Sgspatial International symposium on Advances in Geographic Information Systems, Acm-Gis 2010. Proceedings (pp. 270–279). San Jose, CA, USA: DBLP.

Yasar, H., Hatipoglu, R. S., & Ceylan, M. (2015). An automatic system of detecting changes in aerial images using ANN based contourlet transform. International conference on recent advances in space technologies (pp.199–205). Istanbul, Turkey. doi:10.1109/RAST.2015.7208341

Zhang, N. (2008). Research on key techniques of content-based optical remote sensing image retrieval. PhD diss., National University of Defense Technology.

Zhao, L., Tang, J., Yu, X., Li, Y., Mi, S., & Zhang, C. (2012). Content-based remote sensing image retrieval using image multi-feature combination and svm-based relevance feedback. Lecture Notes in Electrical Engineering, 124, 761–767.

Zhou, W., Newsam, S., Li, C., & Shao, Z. (2017). Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval. Remote Sensing, 9(3), 489.