Journal Pre-proof

Application of image retrieval based on convolutional neural networks and Hu invariant moment algorithm in computer telecommunications

Zhuang Wu, Shanshan Jiang, Xiaolei Zhou, Yuanyuan Wang, Yuanyuan Zuo, Zhewei Wu, Lei Liang, Qi Liu

PII: S0140-3664(19)31247-2
DOI: https://doi.org/10.1016/j.comcom.2019.11.053
Reference: COMCOM 6055

To appear in: *Computer Communications*

Received date : 24 September 2019
Revised date : 20 November 2019
Accepted date : 30 November 2019

Please cite this article as: Z. Wu, S. Jiang, X. Zhou et al., Application of image retrieval based on convolutional neural networks and Hu invariant moment algorithm in computer telecommunications, *Computer Communications* (2019), doi: https://doi.org/10.1016/j.comcom.2019.11.053.

Application of Image Retrieval Based on Convolutional Neural Networks and Hu Invariant Moment Algorithm in Computer Telecommunications

Zhuang Wu[1,2*], Shanshan Jiang[1], Xiaolei Zhou[1], Yuanyuan Wang[1], Yuanyuan Zuo[1], Zhewei Wu[1,2], Lei Liang[1], Qi Liu[1]

[1]School of Management and Engineering, Capital University of Economics and Business, Beijing 100070, China
[2]CTSC center, School of Management and Engineering, Capital University of Economics and Business, Beijing 100070, China

**Corresponding author(E-mail: wuz9080@163.com)*

Abstract: Image recognition and retrieval is an important application field of digital image and it has a wide range of application scenarios in the computer telecommunications. Convolutional neural networks (CNN) is a kind of feed-forward neural networks which includes convolution calculation and which has a deep structure and it is one of the typical algorithms of deep learning. In recent years, it has become a highly efficient recognition algorithm which has been widely applied in such fields as pattern recognition and image processing. Its characteristics include few training parameters and strong adaptability. On the other hand, Hu invariant moment algorithm-a conventional algorithm, is also extensively used in various image processing fields due to its simple calculation and high efficiency. This paper analyzes the research background and significance of both CNN and Hu invariant moment algorithm, and introduces their research status. Besides, it also analyzes the results of color-, distance- and weight-based Hu invariant moment algorithm, and compares it with CNN to serve as a theoretical support for better achieving image classification, recognition and retrieval technology.

Keywords: Convolutional Neural Networks; Hu Invariant Moment; Image Recognition and Retrieval.

1 Introduction

Image feature extraction and analysis are significant research contents of computer vision and they are widely used in the computer telecommunications including object recognition and image retrieval. Floating-point features take up enormous storage space and much calculation cost and they cannot meet the demands of increasingly large-scale image data processing [1]. Convolutional neural networks (CNN) can be dated back to as early as the 1960s when Hubel and Wiesel were conducting the study of visual information processing of cats. In their study, they had discovered a kind of cell which has complex structures. This cell would respond to the local regions in visual information. Soon afterwards, they had brought up the concept of perception field. On this basis, Fukushima in 1984, had proposed Neocognitron, which by adopting the structure of alternating local feature extraction layer and feature transform layer, makes the model maintain the ability for feature recognition when displacement or distortion occurs to objects[2]. Though Neocognitron model does not perform supervised learning by means of back propagation (BP), it is still considered as the first realization of CNN. Later on, Y.LeCun and others had designed the CNN model by using BP method based on Neocognitron and plenty of subsequent

CNN models are its improved versions. Because of its simple calculation and high efficiency[3]. Hu invariant moment algorithm has far-ranging application prospects. In the construction of Hu invariant moment, remove the impact of image translation through use of central moment, eliminate the influence of image scaling via normalized processing and achieve the property of rotation invariance by constructing polynomial. Moments of different orders reflect different properties of the target. Low order moments reflect the basic shape while high order moments show the details and complexity of the target. Low-order Hu invariant moment can better reflect the shape features of the target, but in order to describe target features more meticulously, higher order invariant moment features are needed. However, conventional image retrieval methods are quite sensitive to noises and their quantization process is so unduly simple that it loses much detail information, lacks sufficient judgment ability and needs to be further improved in terms of their performance[4]. This paper takes as the research subjects CNN and Hu invariant moment algorithm both of which have learning ability and conducts comparison experiment on the cases of image retrieval. Image retrieval in essence, is a kind of approximate matching, i.e. the distance between two images is calculated via similarity algorithm. An image retrieval process is hoping to find an image which is the same as or the most similar to the user query image and it extracts image features by using artificial tagging or deep learning algorithm, then compares with the features in the sample database, and returns to the user the image which is the most similar to the query features. The two commonly-seen steps of image retrieval are image feature extraction and feature matching. Usually, image feature extraction is divided into two kinds: low-level features and semantic contents. Low-level vision includes color, texture and shape while semantic contents require to recognize and explain the objects, extract the high-level semantic of the image, and perform matching with the image features extracted[5][6]. An image is normally made up of high-dimensional vectors. Among mass image dataset, it takes much time to compare the similarities among these vectors; so it is not applicable for practical engineering. Image feature matching is the 2nd step of image retrieval and its purpose is to match the image feature extracted with the image. Generally, two images can accurately find the same or the similar images by calculating the similarity with measure algorithm. For example, through the result calculated by Euclidean distance, the smaller the distance, the more similar the images while the large the distance, the more dissimilar the images. It is worth noting that different features require different measure algorithms in similarity calculation. After image retrieval is done, it needs to evaluate the retrieval result and for different image features, different evaluation methods shall be adopted. To find a proper evaluation criterion is an essential step in image retrieval. Totally there are three kinds of evaluation criteria: precision, recall and mean average precision[7]. As precision and recall mutually restrict each other, the concept of average precision (AP) is introduced. Its result calculates the corresponding areas to the curves of precision and recall. Because images are multi-class, multi-class mean are taken, i.e. mean average precision (MAP).

The special contributions of this paper include:

●it analyzes deep CNN and studies the impact of its characteristics and parameter changes.

●it analyzes and studies the results of color-, distance- and weight-based Hu invariant moments.

●it studies the shape-based image retrieval technique, builds a library for intelligent retrieval by extracting image features.

● it makes comparison experiments to CNN and Hu invariant moment algorithm and analyzes their respective characteristics and scopes of application according to the experiment result.

The rest of this paper is organized as follows. Section 2 discusses related work, followed by the related theory including Convolutional neural network and Hu invariant moment algorithm in Section 3. Section 4 shows the simulation experimental results and analysis, and Section 5 concludes the paper with summary and future research directions.

2 Related Work

The study of computer vision has the purpose of making computer cognize the world like humans and have self-discrimination and comprehension ability. Image retrieval can be divided into two types: one based on boundary, which only makes use of the outer contour of the shape without taking into consideration the internal features of the shape, and the other based on region, which uses the entire regional features of the shape. Hu invariant moment represents the shape of the image. As the representation of image features with moments has such spatial geometric invariance as rotation, translation and scaling invariance, it has been widely used to describe shape features. Image feature extraction mainly includes color feature extraction, texture feature extraction, shape feature extraction and spatial relationship feature extraction. These methods can automatically or basically automatically extract features from the image and describe the properties of a certain aspect of the image. In some special application fields, features can also be extracted by means of priori knowledge. Image feature is the quantification of visual effect of the image. Good features can better represent people's visual feelings of the image. Image features can be divided into low-level visual features and high-level semantic features. Although the ideal query of most users are based on high-level semantic features, we can only use low-level features for effective query with the existing technology conditions, especially under the constraint in which the image library has no priori knowledge. The extraction algorithm of low-level feature is the content-based image retrieval technology and it is one of the techniques which have been firstly studied. Now the research direction mainly focuses on the image retrieval of colors, textures, shapes and spatial relationship. Image feature extraction and analysis are the fundamental research contents of computation vision and they are widely applied in object recognition and detection, object tracking, image retrieval, texture classification, three-dimensional reconstruction and so on. Content-based image retrieval (CBIR) has always been an important research topic in computer vision. In early 1990s, researchers had studied and explored CBIR tasks by using the methods of global features, local features and convolutional features of image successively and made remarkable achievements. With its extraordinary performance in image feature recognition, geometric moment method has always been widely studied. Recently, CNN-based image representation method has attracted more and more attention and this method has also miraculous performance[8].

Convolutional neural network and Hu invariant moment algorithm have been extensively applied in the following fields: two-dimensional image processing, pattern recognition, machine vision and shape recognition and the like. Besides, they can also better sole the questions in various fields. As one artificial neural network, CNN is proposed based on Neocognitron and to solve pattern recognition and it has become a research focus of current speech analysis and image recognition. This network is a multi-layer hierarchical neural network. In every layer, the neurons are of the same type and they can be simple, or complex or hyper-complex neurons. Besides, between layers, there is very rare connection with fixed pattern. After introducing the basic CNN structure and its neuron model, it continues to discuss the training process of CNN. When the necessary features have been pre-set, supervised algorithm is adopted and the network learns layer by layer; otherwise, unsupervised learning is conducted. Its weight-sharing network structure makes it more similar to biological neural network and it lowers the complexity of the network model and reduces the number of weights. This advantage is more significant when the network input is multi-dimensional image: it takes the image directly as the network input and avoids the complex feature extraction and data reconstruction process of conventional recognition algorithms. Convolutional network is a multi-layer perceptron specially designed to recognize two-dimensional shape. The structure of this kind of network is highly invariant to translation, scaling, slanting or other forms of distortion[9].

Hu M.K had used normalized 2nd- and 3rd-order central moments to construct 7 Hu invariant moments according to the theory of algebraic invariant. In the construction of Hu invariant moments, he removes the impact of image translation through the use of central moment, eliminates the influence of image scaling via normalized processing and achieves the property of rotation invariance by constructing polynomial. Moment is the operator to describe image features and it has significant applications in pattern recognition and image analysis. So far, there are following common moment descriptors: geometric moment, orthogonal moment, complex moment and rotational moment. Among them, geometric moment was proposed the earliest, it has a simple form, and it is the most fully studied. Geometric moment has certain description ability in simple image. Although it does not work as well as the other three kinds of moments with regards to discrimination, it is quite simple and it usually needs a figure in its representation compared with other moments. Geometric moment was proposed by Hu (Visual pattern recognition by moment invariants) in 1962 and its main idea is to use several region-based transform-insensitive as shape features. Moment feature mainly represent the geometric feature of image regions, also known as geometric moment. Due to its features of rotation, translation and scale invariant, it is also referred to as invariant moment. In image processing, geometric invariant moment can represent object as an important feature and such operations as image classification can be conducted according to this feature. Moment is used to reflect the distribution status of stochastic variables in statistics and when promoted in dynamics, it is used to depict the mass distribution of space object. Likewise, if we consider the gray value of the image as a two- or three-dimensional density distribution function, then moment method can be used in image analysis and image feature extraction. Most frequently, the zero order of the object represents the “mass” of the image[10] [11].

3 Related Theoretical Bases

3.1 Hu Invariant Moment Algorithm

In CBIR, comparison shall be conducted on the similarities between query image features and the features in sample image library and then the result is given back to the user by a decreasing order of similarities; therefore, the effectiveness and accuracy of the image greatly affect the query result. One anticipant image retrieval algorithm shall meet the following two properties: (1) it can accurately describe characteristics of the image; (2) it has low calculation costs. Therefore, efficient image retrieval algorithm ensure real-time operation of the machine, which is also very important to retrieve large corpora and the image or detection object in mobile devices. Hu invariant moment can reduce the pre-processing steps such as size normalization and position centering of the image to be detected, cut the operation cost, and improve the operation speed. As local invariant features are used, it makes object recognition and detection better handle these complicated circumstances. It is a key research point of current computer vision and it is mainly used to describe the positions and relationship of pixels in an image as well as the feature description of objects, including image edge, contour and corner and then combine and transform these features before forming stationary feature descriptors which are easy to match. In order to facilitate the calculation of similarities of feature vectors in image retrieval, binary descriptor is used to represent image features, i.e. a string of binary values of 0 and 1. The practice shows that directly using origin moment or central moment as image feature cannot ensure that the features are translation, rotation and scale invariant. As a matter of fact, if only the central moment is used to represent image feature, then the feature is only translation invariant. If normalized central moment is used, then the feature is not only translation invariant, but also scale invariant[12] [13].

Hu moment uses 2nd- and 3rd-order central moments and constructs 7 invariant moments, which can help describe the region shape of an image and has rotation, translation and scaling invariance. Invariant moment refers to the moment feature which remains the same after translation, rotation and scaling. Hu invariant moment is a region-based image shape description method. Normalize the central moments of the rest orders of a 2D discrete image and get the normalized central moments. The idea of color moments is that color distribution of the image can be represented with its moments. As the color distribution information is mainly in low-order color moments, the color features of an image can be displayed by only using the 1st-, 2nd- and 3rd-order central moments of colors, which indicate the average color, standard variance and cube root asymmetry respectively. Feature extraction mainly extracts low-level features of the image, including colors, textures, shape and spatial relationship and among others.

In continuous circumstances, the image function is $f(x,y)$, then the $p+q$th -order geometric moment (standard moment) of the image is defined as

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dxdy \tag{1}$$

$p+q$th -order central moment is defined as

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\overline{x})^p f(y-\overline{y})^q f(x,y)dxdy \qquad p,q=0,1,2\ldots \tag{2}$$

where $\overline{x}$ and $\overline{y}$ represent the centers of gravity of the images.

$$\overline{x} = m_{10}/m_{00}, \qquad \overline{y} = m_{01}/m_{00} \tag{3}$$

For discrete digital image, replace integral with sum.

$$m_{pq} = \sum_{y=1}^{N}\sum_{x=1}^{M} x^p y^q f(x,y) \qquad p,q=0,1,2\ldots \tag{4}$$

$$\mu_{pq} = \sum_{y=1}^{N}\sum_{x=1}^{M} (x-\overline{x})^p (y-\overline{y})^q f(x,y) \qquad p,q=0,1,2\ldots \tag{5}$$

N and M are the height and width of the image respectively.

The normalized central moment is defined as

$$\eta_{pq} = \mu_{pq} / (\mu_{00}{}^{\rho}) \tag{6}$$

where $\rho = (p+q)/2+1$. (7)

Use second- and third-order normalized central moments to construct 7 invariant moments $M1 \sim M7$:

$$\begin{aligned}
&M_1 = \eta_{20} + \eta_{02} \\
&M_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11} \\
&M_3 = (\eta_{30} + 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
&M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
&M_5 = (\eta_{03} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + 3\eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
&(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
&M_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12}) + (\eta_{21} + \eta_{03}) \\
&M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
&(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned} \tag{8}$$

To recognize the image through the features formed by Hu moments is very fast, but it has a low recognition rate. Usually, Hu invariant moments are used to recognize large objects in the image and they can better describe image shape and the image texture feature cannot be too complex [14].

3.2 Convolutional Neural Network

A typical convolutional neural network is comprised of several convolutional layers, pooling layers and fully-connected layers. The convolutional layer and pooling layer usually alternate in showing up in the front part of the network while the rear part is fully-connected layer. Convolutional layer produces feature mapping through convolution in the input image. Pooling layer uses the sliding window of definition and takes the maximum or the mean value as the value to be passed on to the next layer. Before conducting down-sampling, the output of convolutional layer also needs to be delivered to a non-linear function, which is called as activation function and which can perform non-linear transformation on data. To predict the input data, connect the output fraction of the final fully-connected layer of CNN to the loss function, which can normalize the fraction to the probability distribution of every tag. In other words, the bigger the fraction, the higher probable for the input image to belong to that tag. In the end, minimize the loss function between the tags of predicted value and real value and optimize the network parameters through regularization terms[15] [16].

Convolutional Neural Networks (CNN) is better than fully-connected network with regards to time and memory. CNN is derived from fully-connected network. As every neuron in the ith layer of fully-connected network is connected with all neurons in the (i-1)th layer, namely that every neuron in the hidden layer is connected with all the neurons in the input layer, every connection has its corresponding parameter (weight). As the size of the input image and the number of hidden layers increase, the network parameters will also increase greatly. Even for small fully-connected network, the reason why the number of network parameters is very big is that every connection of the neurons between layers has different parameters. Therefore, it can be considered that a group of neurons are provided with the same parameter and the neurons within a group of neurons will be allocated with the same parameter; in this way, the number of network parameters will reduce greatly. If fewer pixels can be used to get the same or better result, the number of parameters will be greatly reduced. Therefore, network parameters can be optimized from this perspective. Usually in the image analysis, input image is transformed into pixel matrix, every pixel of which is highly correlated to its surrounding pixels. The bigger distance between two pixels, the more unrelated they are. For example, facial pixels are related to the pixels around the face, but they are barely related to the pixels of sky and ground. As CNN uses shared weight and fewer parameters, CNN structure usually has more layers, which is its unique characteristic which the fully-connected network doesn't have.

CNN has several internal layers and it is a neural network model with the pipeline method which has several processing phases. In the experiment of this paper, the image enters into the neural network from one end and then the accuracy rate of image recognition is output from the other end. There are three types of layers in CNN: convolutional layer, pooling layer and fully-connected layer. The structure of CNN is shown as Fig.1. This is also a neural network of deep learning as it also includes several internal layers, but now these internal layers have one of the above three types.

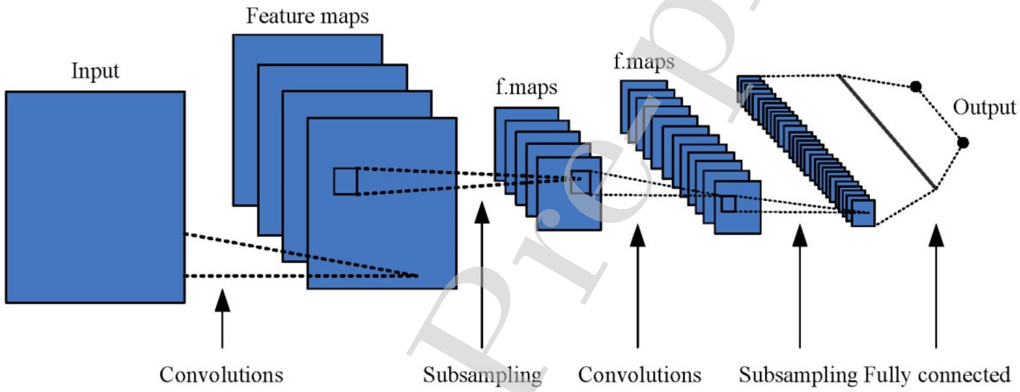


Figure.1 Structure of convolutional neural network

As many intermediate layers as possible can be set in the neural network. The neurons in the neural network can be represented with the formula below.

$$y = \sigma(wx + b) \tag{9}$$

where w is the weight, b is the bias, and $\sigma()$ is the non-linear function, also known as the activation function, which calculates the neuron input $wx + b$. Activation function has several optional forms[17].

CNN scans input data by means of mask. Every input of the mask needs to go through an activation function. For a given mask, the activation function is the same, which can reduce the number of weights. Fig.2 is the convolution matrix.

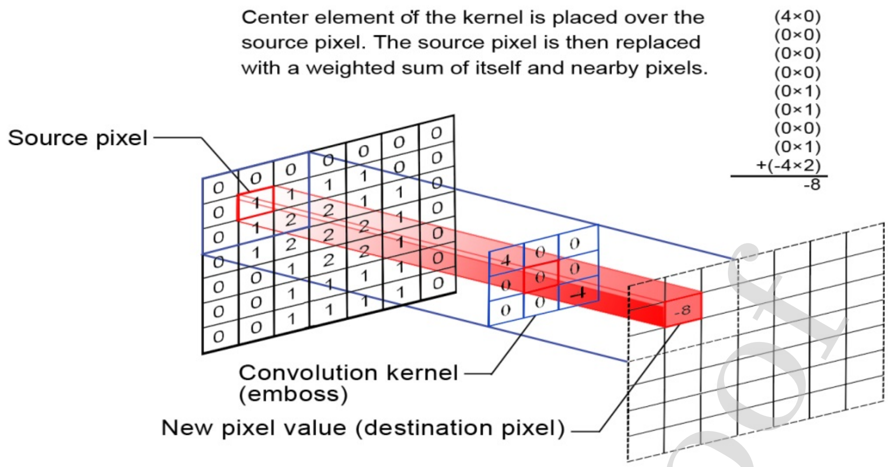


Figure.2 Convolution matrix

Pooling layers obtains the subset of the output result of convolutional layer and it continues to convey it to the subsequent layer. Pooling layer has no weights and it can take the maximum value, or the mid-value or the mean value of the elements in the pool as the output. Pooling function can take these pooling methods for users to choose. In the realization process, pooling function divides the input into n*n sub-regions and return to a corresponding matrix.

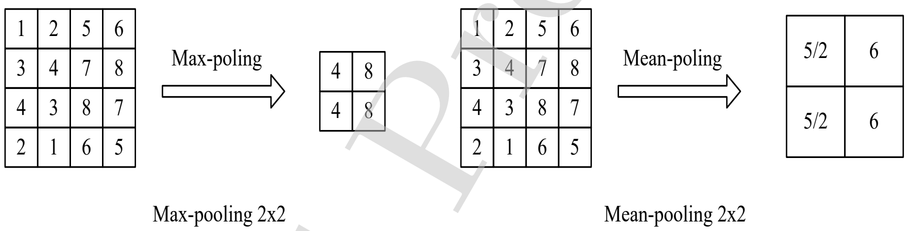


Figure.3 Pooling process with a pooling step-length of 2

CNN only trains one layer of the network. It fixes the previous $k-1$ layers, takes its output as the input and starts to train the kth layer. The training of every layer can be supervised. For example, take the classification error of every step as the object function. But more frequently, unsupervised method is used. The weights of separate trainings of every layer are used to initialize the weight of the final deep network. In the end, “slightly adjust” the entire network, in another word, put together all layers to optimize the training errors of the tagged training sets[18] [19].

4 Test Experiment and Analysis

With the development of Internet information technology, how to effectively save and retrieve mass image data has increasingly attracted people’s attention. Therefore, it is of great importance to effectively build image database, construct image retrieval engine and use the key image data. This paper studies the shape-based image retrieval technique and conducts intelligent retrieval by extracting image features and building database. This case chooses to conduct image retrieval with convolutional neural network and Hu invariant moments as the standard.

4.1 Dataset

This paper chooses CIFAR-10 dataset which is widely applied in the field of image classification. This dataset has tag and comes from a larger-scale dataset. CIFAR-10 dataset totally has 60000 32*32 color images. They are divided into 10 classes: plane, automobile, bird, cat, deer, dog, frog, horse, ship and truck and every class every 6000 images. In this paper, 35 images are stochastically taken from every class in the test and then it makes comparison between convolutional neural network and Hu invariant moment algorithm.

4.2 Difference Calculation of Hu Invariant Moment

Hu invariant moment can calculate the following difference values: (1) color difference of Hu, (2) distance difference of Hu, and (3) difference value of weighted integration of (1) and (2). Image difference calculation can extract the image files which are similar to the test images in the image library and it is a key step of image retrieval. The smaller the difference, the more similar the images. The specific steps are as follows:

Step 1: Conduct edge detection on the images to be retrieved with edge detection operator. Identify the edge detection operator, perform edge extraction on the image and obtain the edge image. After identifying the edge image, conduct contour tracing and obtain outer contour image. The paper chooses automobile and dog in CIFAR-10 dataset as the experiment objects this time and it stochastically selects an image of these two classes of images for experiment.

Step 2: Use the 7 invariant moments of Hu as the shape feature vectors. Confirm the object region image, calculate the 7 Hu invariant moments of the object region and build it into the shape feature vector of this image. Determine the shape feature vector, perform internal normalization and save the feature value in the image feature library.

Step 3: Conduct image similarity matching. Calculate Hu invariant moments of the image to be retrieved and conduct image similarity matching according to the recommendation index.

(1) Hu-based Color Difference Calculation

Color difference is a focus of chromatology and it quantizes a concept. Before quantization, people can only use adjectives to roughly describe this concept, which makes it rather inconvenient for the workers who have strict with colors. Color difference can be simply calculated through the Euclidean distance in color space or via the complex and uniform human intuition formula of CIE. Here, perform multi-level quantization on the color matrix of the image, conduct weighting, calculate the histogram and perform normalization processing, and obtain the final color difference. According to the experiment result, see the color difference calculate result in Tab.1 below.

Table.1 Comparison of recommended accuracy of color difference of every class

Class	Automobile	Dog	Plane	Cat	Bird	Frog	Ship	Deer	Truck	Horse
Accuracy	37.5%	43.75%	12.5%	18.75%	12.5%	18.75%	12.5%	31.25%	43.75%	43.75%

It can be seen that after testing, the average recognition accuracy of automobile, dog, plane, cat, bird, frog, ship, deer, truck and horse are 37.5%, 43.75%, 12.5%, 18.75%, 12.5%, 18.75%, 12.5%, 31.25%, 43.75%, and 43.75% respectively.

(2) Hu-based Distance Difference Calculation

The performance of image retrieval not only relies on the image features extracted. After such image features as color, texture and shape are extracted and indexes are built, the key of image retrieval lies in the similarity measure (or distance measure) function. It directly affects the result and efficiency of image retrieval. This paper uses Hu-based distance difference calculation based on selecting Hu invariant moments and the calculation result is shown in Tab.2.

Table.2 Comparison of Recommended Accuracy of Distance Differences of Every Class

Class	Automobile	Dog	Plane	Cat	Bird	Frog	Ship	Deer	Truck	Horse
Accuracy	56.25%	37.5%	18.75%	18.75%	25%	18.7%	25%	31.25%	43.75%	43.75%

It can be concluded that after testing, the recognition accuracy of automobile, dog, plane, cat, bird, frog, ship, deer, truck and horse on average are 56.25%, 37.5%, 18.75%, 18.75%, 25%, 18.75%, 25%, 31.25%, 43.75%, and 43.75% respectively.

(3) Weighted integration

As one method has certain flaws in accuracy, this paper makes some improvements based on the above methods, adds together the two difference calculation results according to the corresponding weight calculated and obtains the final result. The calculation result is described in Tab.3 below.

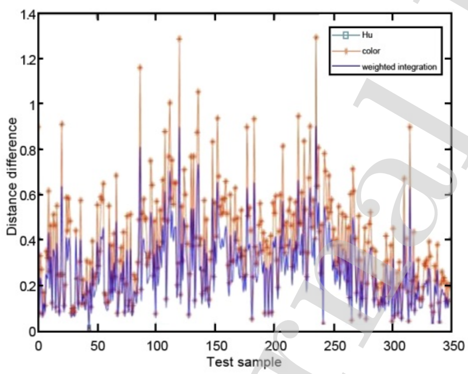
Table.3 Comparison of Recommended Accuracy of Two Difference Weights of Every Class

Class	Automobile	Dog	Plane	Cat	Bird	Frog	Ship	Deer	Truck	Horse
Accuracy	43.75%	37.5%	18.75%	18.75%	12.5%	25%	18.75%	31.25%	43.75%	43.75%

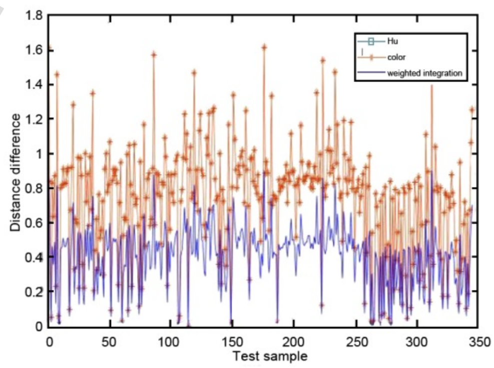
After the test, it is clear that the average recognition accuracy are 43.75% (automobile), 37.5% (dog), 18.75% (plane), 18.75% (cat), 12.5% (bird), 25% (frog), 18.75% (ship), 31.25% (deer), 43.75% (truck), and 43.75% (horse) respectively.

(4) Comparison of Three Methods

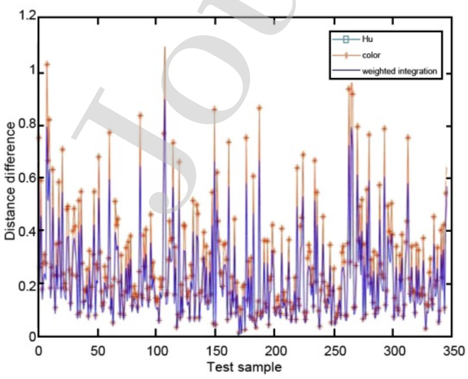
According to the above three calculation methods, this paper compares the recommended indexes for a more intuitive understanding of them. Please find the result description below.

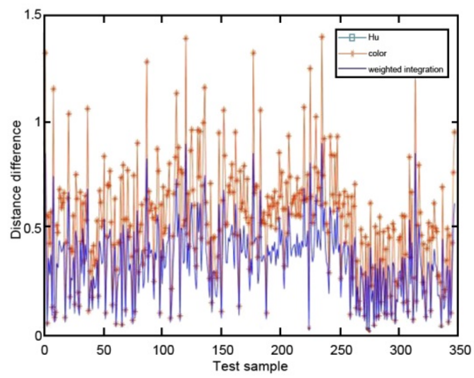


(a)Automobile

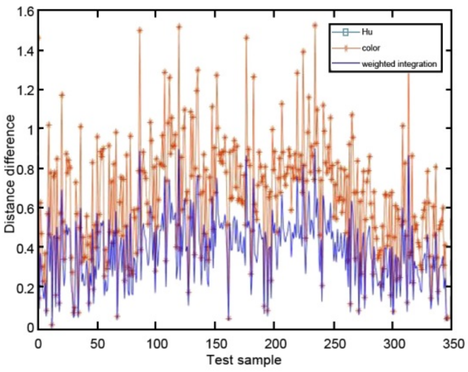


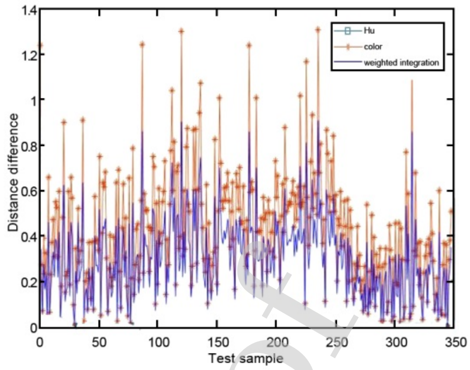
(b) Bird



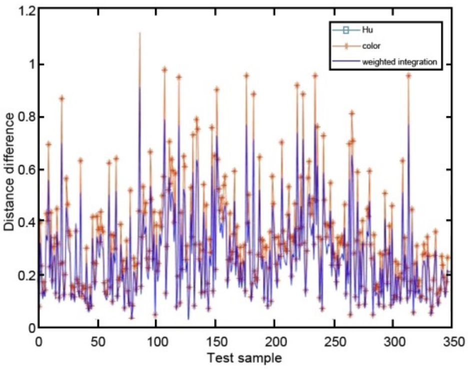


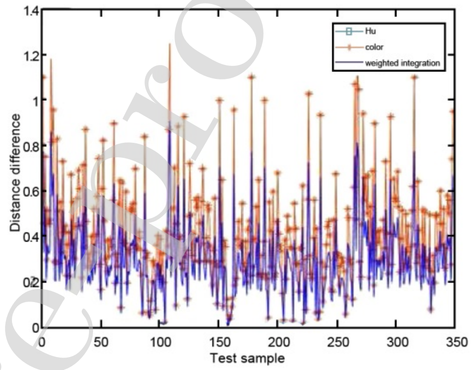
(c)Dog (d) Ship



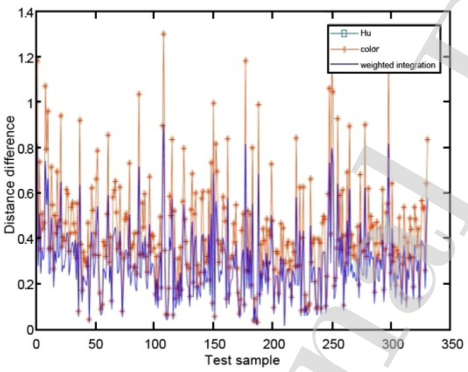


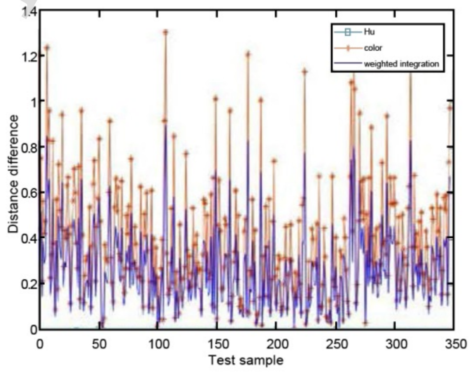
(e) Plane (f) Truck





(g) Cat (h) Deer





(i) Frog (j) Horse

Figure.4 Comparison of Recommended Indexes of Three Methods

As shown above, in the five classes: automobile, dog, truck, horse and deer, these three methods are positively correlated and their recommended indexes are strongly related while in the classes of cat, bird, frog, plane and ship, the methods are badly correlated and their recommended indexes are unrelated. Therefore, in case of higher accuracy, the recommended indexes of three methods are positively correlated and with lower accuracy, they are unrelated. Among them, the recommended index calculated with weight is more practical.

Retrieve the closest 16 images in the image library as the retrieval results. See Fig.5 below for the retrieval results.

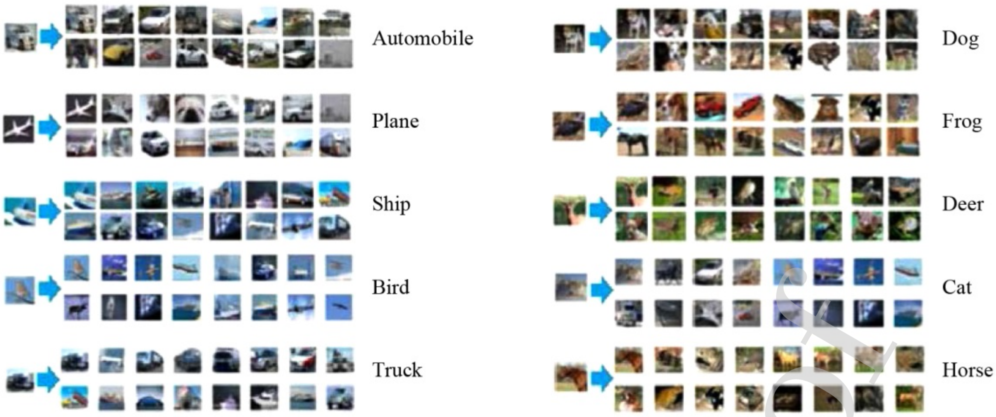


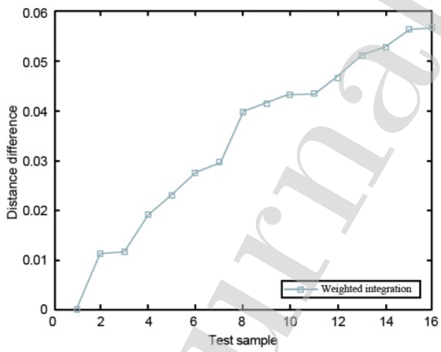
Figure.5 Recognition results

After being tested, the average recognition accuracy for automobile is 54.53%, dog 43.75%, plane 18.75%, cat 20.32%, bird 18.75%, frog 25%, ship 18.75%, deer 28.13%, truck 56.25% and horse 50%.

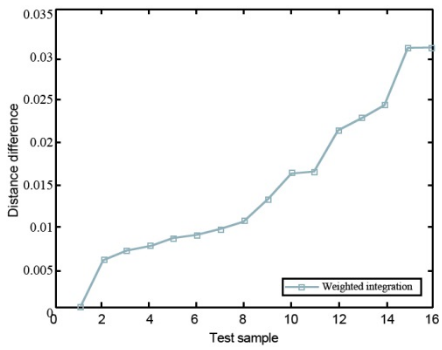
Table.4 Recognition accuracy

Class	Automobile	Dog	Plane	Cat	Bird	Frog	Ship	Deer	Truck	Horse
Accuracy	54.53%	43.75%	18.75%	20.32%	12.5%	25%	18.75%	28.13%	56.25%	50%

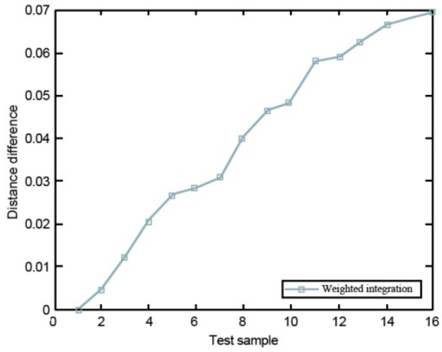
Compared with automobile and dog, plane has a more special shape and the huge fuselage has brought certain difficulties for image retrieval. Besides, as plane can only be photographed in the runway or the sky, the external environment usually increases image noises and thus makes it more difficult for image retrieval. But image pre-processing still has certain effect. For example, in automobile retrieval, the pre-processed automobile image in haze has been retrieved. The retrieval result similarity of every class is shown in Fig.6 below.



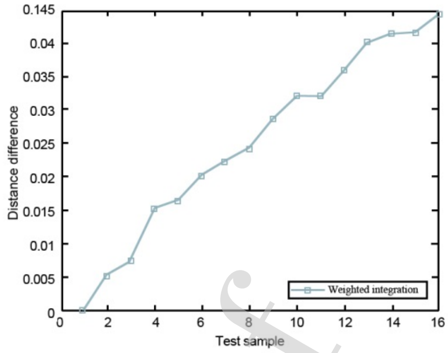
(a) Result similarity of automobile



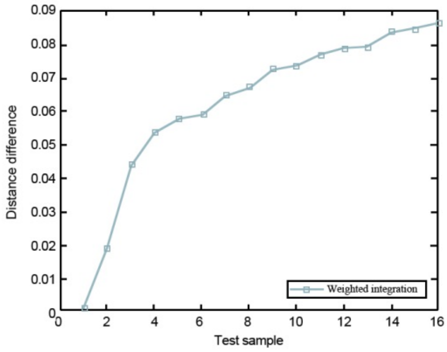
(b) Result similarity of bird



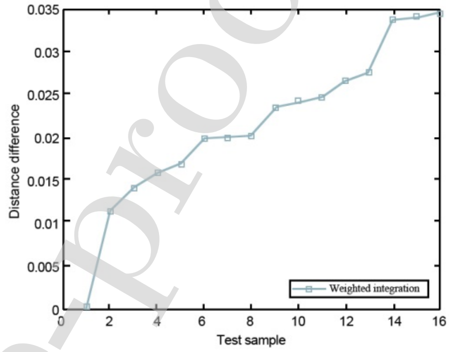
(c) Result similarity of dog



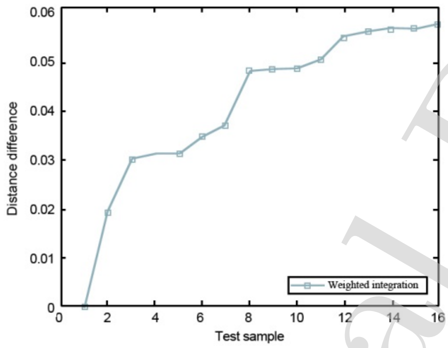
(d) Result similarity of ship



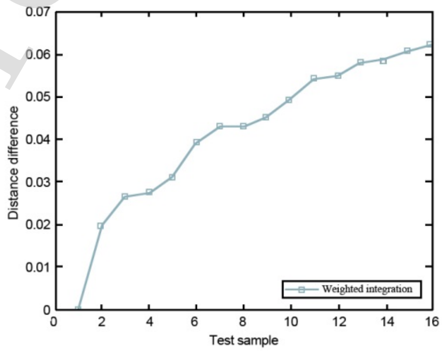
(e) Result similarity of plane



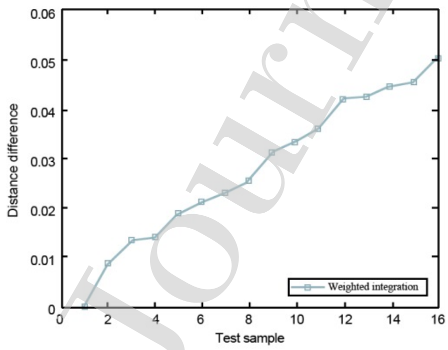
(f) Result similarity of truck



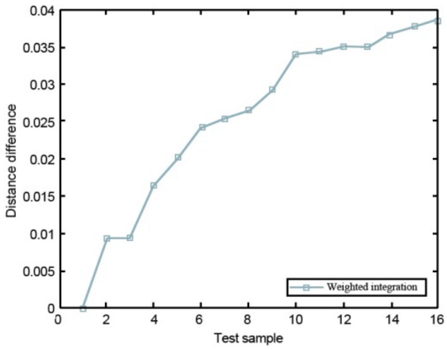
(g) Result similarity of cat



(h) Result similarity of frog



(i) Result similarity of deer



(j) Result similarity of horse

Figure.6 Retrieval result similarity of every class

The above experiment result demonstrates that to conduct image retrieval with Hu invariant moments as features have higher implementation efficiency and effective retrieval result. For different images, calculate the feature vectors of Hu, compare it with the original image library and extract the result image as the output to reflect the process of image retrieval after a series of pre-processing and it has certain practical meaning.

4.3 Analysis of Hu Invariant Moment Algorithm in Image Retrieval

Through the above comparison and experiment, it can be seen that for different classes, Hu invariant moments have different retrieval accuracy. Why the accuracy of horse is as high as 50% and that of bird is only 12.5%. This paper conducts research and exploration from the perspective of image features. The principle to be followed by image feature extraction is: the image features shall be slightly different for the same class and greatly different for different classes.

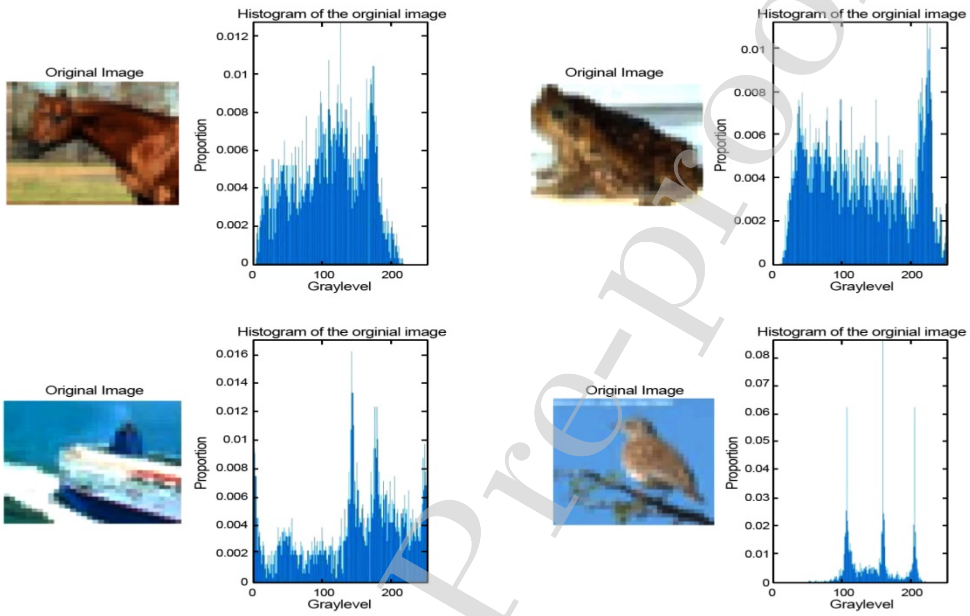


Figure.7 Comparison of feature distribution of dataset

So, it can be easily found from the comparison of horse, frog, cat, ship, and bird that the more obvious the image features, the better the classification result and the less significant the image features, the worse the classification result. Horse has obvious image features and its gray values are concentrated in a small distribution ratio. And bird has the least obvious image features, which are scattered and have small numerical values, and it is not good for retrieval. It can be learnt from the experiment result that many images are no clear in our real life; so for now, essential image processing technique and more in-depth exploration in image retrieval are still very important.

4.4 Comparison between Convolutional Neural Network and Hu Invariant Moment in Image Retrieval

Table.5 Retrieval Accuracy of Convolutional Neural Network

Class	Plane	Automobile	Bird	Cat	Deer	Dog	Ship
Accuracy	41%	56%	37%	53%	45%	51%	39%

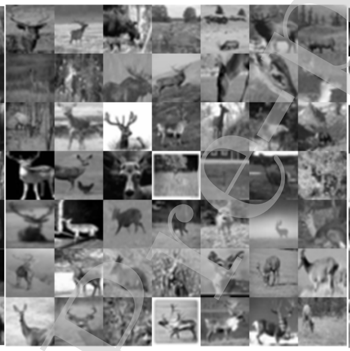
Test with convolutional neural network and the result shows that the recognition accuracy for plane is 41%, automobile 56%, bird 37%, cat 53%, deer 45%, dog 51%, and ship 39%. With cat as example, the recognition accuracy for cat with Hu invariant moment in image retrieval is 20.32% while the figure rises to 53% in convolutional neural network. Compared with Hu invariant moment algorithm, convolutional neural network has larger computation and a low speed. Its learning process has ensured the accuracy of the result though, it also increased the computation burden. On the contrary, Hu invariant moment algorithm can give a result within seconds. Therefore, for the images with high image feature values, Hu invariant moment algorithm is recommended while those images with low values, convolutional neural network is recommended.


(a) Plane


(b) Automobile


(c) Bird


(d) Cat


(e) Deer


(f) Dog


(g) Ship

Figure.8 Image retrieval result of convolutional neural network

The convolutional network has the following advantages in image processing compared with the general neural network: The input image and the topology of the network can be well matched, the feature extraction and pattern classification are performed and generated in the training simultaneously, the weight sharing can reduce the network training. The parameters make the neural network structure simpler and more adaptable. We will further study retrieving images of lower quality or images with more similarities for the future work.

5 Conclusion

Image acquisition, analysis, retrieval and recognition are greatly important research topic in the computer telecommunications. Computer vision is closely related to digital image and it uses computer to analyze the image or video collected by computer to simulate human visual system and complete corresponding processing and interpretation. Convolutional Neural Networks (CNN) has already become a research hotspot of such fields as image, text and speed recognition. Hu invariant moment algorithm is one of the conventional image recognition and retrieval methods. Moment is a statistical form of image and its calculation needs to use all related pixel points in the image or the region, therefore, the global property of the object is described from the global perspective. This paper uses CNN and Hu invariant moment to extract image features, conducts test experiment through different methods, and obtains the comparison result. Compared with Hu invariant moment algorithm, CNN has higher accuracy, but the calculation is huge and the speed is slow while using Hu invariant moment algorithm can improve efficiency of feature extraction by avoiding mass data from independently learning image features.

Acknowledgements

This work was supported by the National Social Science Foundation of China (No.16BGL145).

Conflict of interest statement

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

References

[1] Thomas Fischer. "PCA-based Supervised Identification of Biological Soil Crusts in Multispectral Images", MethodsX, 6, (2019), pp.764-772.

[2] Jie Shao, Yongsheng Qian. "Three Convolutional Neural Network Models for Facial Expression Recognition in the Wild", Neurocomputing, 355(25), (2019), pp.82-92.

[3] Boukaye Boubacar Traore, Bernard Kamsu-Foguem, Fana Tangara. "Deep convolution neural network for image recognition", Ecological Informatics, 48(11), (2018), pp.257-268.

[4] Imad Batioua, Rachid Benouini, Khalid Zenkouar, Azeddine Zahi. "Image Classification using Separable Invariants Moments Based on Racah Polynomials", Procedia Computer Science, 127, (2018), pp. 320-327.

[5] Wissam A. Albukhanajer, Yaochu Jin, Johann A. Briffa. "Classifier Ensembles for Image Identification using Multi-Objective Pareto Features", Neurocomputing, 238(17), (2017), pp.316-327.

[6] Xiaolin Bi, Sifan Lin, Siqi Zhu, Hao Yin, Zhenqiang Chen. "Species Identification and Survival Competition Analysis of Microalgae via Hyperspectral Microscopic Images", Optik, 176(1), (2019), pp. 191-197.

[7] Anh Viet Phan, Minh Le Nguyen, Yen Lam Hoang Nguyen, Lam Thu Bui. "DGCNN: A Convolutional Neural Network over Large-Scale Labeled Graphs", Neural Networks, 108(12), (2018), pp.533-543.

[8] Na Huang, Jingsha He, Nafei Zhu, Xinggang Xuan, Chengyue Chang. "Identification of The Source Camera of Images Based on Convolutional Neural Network", Digital Investigation, 26(9), (2018), pp. 72-80.

[9] Jayme G. A. Barbedo, Guilherme B. Castro. "Influence of Image Quality on The Identification of Psyllids using Convolutional Neural Networks", Biosystems Engineering, 182(6), (2019), pp.151-158.

[10] Yaxin Sun, Guihua Wen, Jiabing Wang. "Weighted Spectral Features Based on Local Hu Moments for Speech Emotion Recognition", Biomedical Signal Processing and Control, 18(4), (2015), pp.80-90.

[11] Abdeslam Hmimid, Mhamed Sayyouri, Hassan Qjidaa. "Fast Computation of Separable Two-Dimensional Discrete Invariant Moments for Image Classification", Pattern Recognition, 48(2), (2015), pp.509-521.

[12] Chandan Singh, Sukhjeet K. Ranade, Karamjeet Singh. "Invariant Moments and Transform-Based Unbiased Nonlocal Means for Denoising of MR Images", Biomedical Signal Processing and Control, 30(9), (2016), pp.13-24.

[13] Luhong Diao, Juan Peng, Junliang Dong, Fanyu Kong. "Moment Invariants under Similarity Transformation", Pattern Recognition, 48(11), (2015), pp.3641-3651.

[14] Huihui Li, Xing Jin, Ning Yang, Zhe Yang. "The Recognition of Landed Aircrafts Based on PCNN Model and Affine Moment Invariants", Pattern Recognition Letters, 51(1), (2015), pp.23-29.

[15] Imad Batioua, Rachid Benouini, Khalid Zenkouar, Azeddine Zahi. "Image Classification using Separable Invariants Moments Based on Racah Polynomials", Procedia Computer Science, 127, (2018), pp. 320-327.

[16] Tong Qiao, Florent Retraint, Rémi Cogranne, Thanh Hai Thai. "Individual Camera Device Identification from JPEG Images", Signal Processing: Image Communication, 52(3), (2017), pp.74-86.

[17] Mindaugas Jurevicius, Jonas Skeivalas, Robertas Urbanavicius. "Analysis of Surface Roughness Parameters Digital Image Identification", Measurement, 56(10), (2014), pp.81-87.

[18] Rui Wang, Wei Li, Liang Zhang. "Blur Image Identification with Ensemble Convolution Neural Networks", Signal Processing, 155(2), (2019), pp.73-82.

[19] Zhenhua Huang, Xin Xu, Juan Ni, Honghao Zhu, and Cheng Wang. Multimodal Representation Learning for Recommendation in Internet of Things. IEEE Internet of Things Journal, (2019), DOI: 10.1109/JIOT.2019.2940709.

Conflict of interest statement

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Dear Editors:

We would like to submit the enclosed manuscript entitled "Application of Image Retrieval Based on Convolutional Neural Networks and Hu Invariant Moment Algorithm in Computer Telecommunications", which we wish to be considered for publication in "Computer Telecommunications". No conflict of interest exits in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

The special contributions of this paper include:

●it analyzes deep CNN and studies the impact of its characteristics and parameter changes.

●it analyzes and studies the results of color-, distance- and weight-based Hu invariant moments.

●it studies the shape-based image retrieval technique, builds a library for intelligent retrieval by extracting image features.

● it makes comparison experiments to CNN and Hu invariant moment algorithm and analyzes their respective characteristics and scopes of application according to the experiment result.

I hope this paper is suitable for "Computer Telecommunications".

We deeply appreciate your consideration of our manuscript, and we look forward to receiving comments from the reviewers. If you have any queries, please don't hesitate to contact me at the address below.

Thank you and best regards.

Yours sincerely,

Zhuang Wu

Corresponding author:
Name: Zhuang Wu
E-mail: wuz9080@163.com