The improved AdaBoost algorithm application in image retrieval

Si Qing Zhang[1,a] Run Zheng[1,a]
[1]Zhengzhou College of Science & Technology, Zhengzhou China 450064
[a]zsqn@tom.com

Abstract—**In the traditional content-based image retrieval system, for a given query image, the number of relevant images in the database are not far outnumber correlation image. Therefore, a number of negative samples and the number of positive sample is unbalanced, the two class classifier traditional lose effectiveness. In this paper, we will present the OCSVM integration method based on improved AdaBoost to solve this problem. Although OCSVM is seen as a strong classifier, in this way, we are still on the training data in AdaBoost weight updating formula was improved so that the AdaBoost can be integrated with OCSVM.**

Keywords — ***Image retrieval ; AdaBoost ; OCSVM integration***

I INTRODUCTION

With the development of digital image technology, image retrieval has become a hot research topic. From the traditional keyword-based image retrieval to widely used content-based image retrieval. Content-based image retrieval mainly utilizes the low-level visual contents for features to retrieval, for example, color, texture, and shape. However, in the process of retrieval, the users always join their subjective perception in the system, which results in production of a semantic gap between the user's query and the low-level visual content. In order to reduce the semantic gap, a new retrieval method is proposed in the dissertation. It needs the users to segment out the region they are interested and compare the similarity of those fragments. Hence, the retrieval system can learn the users' query requirement and reduce the difference between semantics. At the same time, the correlation coefficient is introduced into the color space to find the correlation between the color channels. Therefore, not only the color consistency between the retrieval images and the query image is kept, but also the accuracy precision is also improved.

In order to further learn the user's query semantic, the relevance feedback is introduced into the retrieval process. The process of relevance feedback can be regarded as a classification problem. The relevant images are labeled as positive class, while the irrelevant images are labeled as negative class. However, the number of negative images is usually much more than that of the positive images, which causes the problem of imbalance data classification. The traditional two-class support vector machine (SVM) may lose its effectiveness on tackling the foresaid classification. Therefore, the one-class support vector machine (OCSVM) and the modified AdaBoost based OCSVM ensemble are utilized in the dissertation. Experimental results demonstrate that the utilized ensemble method gets the better performance and higher accuracy.

In recent years, image retrieval has received more and more attention. Especially with the emergence of cable technology image retrieval based on content, the traditional system of image retrieval performance has been greatly improved. The image retrieval based on content is to reduce the labeling cost based on keywords. Moreover, it can automatically extract given query image low-level visual features, such as color, texture and shape. Then the similarity matching the image features of these characteristics and the database.

As everyone knows, the color is an important attribute of image. At present, people have proposed many retrieval system based on color features, such as color histogram, color moments, dominant color descriptor. The advantages of color histogram is that it has a good rotation, translation invariance. However, it has lost the spatial structure information between color bits. Color moment is based on the concept of probability distribution of moment to express the color distribution. Compared with the color histogram, color moments do not need to quantify the color space. Therefore, dimension is low, processing is convenient. The main color descriptor by extracting a plurality of body color in the image to describe color feature. However, it ignores the spatial distribution information of color.

In addition to color, texture is an important feature of.

So far, a lot of texture image retrieval system has been applied on, such as Gabor transform, wavelet transform, co-occurrence matrix etc.. These texture features describe the texture information of the image surface. In fact, when the search image, people should not only consider the low-level visual features of the image, but also consider the subjective perception of the user's query.

In order to make full use of subjective perception of users, Huang and others in the content-based image retrieval based on relevance feedback retrieval mechanism. Users need to mark the retrieval results in relevance feedback, recorded as positive or negative. Therefore, relevance feedback process can be regarded as a classification problem. In the field of machine learning has many classification methods, and some methods in image retrieval based on relevance feedback is applied. For example, the decision tree (DT), K nearest neighbor (KNN), neural network (NN), genetic algorithm (GA) and support vector machine (SVM).

However, in the traditional content based image retrieval, for a given query image, a number of negative cases in the database are often much more than the positive examples. Therefore, a number of positive and negative examples is extremely unbalanced, which makes the two class classifier traditional lose effectiveness. In order to solve this problem, we use a support vector machine (OCSVM). In addition, a plurality of a class of classifiers can also obtain performance than a single classifier better. In order to reduce the influence of noise and prevent overfitting, Albert and David in OCSVM integrated in the weighted bagging strategy.

Recently, we proposed the integration method of OCSVM based on improved AdaBoost is also in order to solve the problem of unbalanced classification data. In addition, we found that the direct integration of a OCSVM bagging or the result of the boosting method does not improve the performance of single OCSVM, because OCSVM is a strong classifier. In this paper, we will present the OCSVM integration method based on improved AdaBoost used in image retrieval, in order to obtain the more ideal retrieval performance.

Ⅱ THE INTEGRATION OF OCSVM BASED ON IMPROVED ADABOOST

AdaBoost is an integrated algorithm commonly used, through the integration of multiple weak classifiers to achieve more desirable results. Because the OCSVM is seen as a strong classifier, therefore, the traditional AdaBoost can not directly be multiple OCSVM classifiers. Then, we improve the AdaBoost in the weight updating formula, so that it can be integrated with OCSVM. Below we briefly describe the OCSVM integration algorithm based on AdaBoost.

When given a set of training set {x I}iN=1, as well as the data of Xi probability in t iteration of the distribution of Dt (I), OCSVM integration algorithm based on improved AdaBoost can be described as the following algorithm 1. In the algorithm 1, constant index added C ∈ (1,10] such that each OCSVM base classifiers are not the same, so you can change the base classifier is more obvious. In this paper we on the value of C is 5, the epsilon Ti ∈ (0,1], so the index of E ε ti is always greater than 1. We can also see from the algorithm, index C increased the diversity of α t becomes more significan.

Ⅲ EXPERIMENTAL RESULT

We conduct experiments on three image databases. The first two database from Corel GalleryMagic, third from Caltech database. The first database consists of 1000 images, there are 10 categories, each category contains 100 images. Image size is 256*384 or 384*256. Are Indians, beaches, mountains, bus, dinosaurs, elephants, horses, flowers, glaciers and Western-style food. Second contains 500 images, a total of 5 categories, they are butterflies, car, the church, the trees and flowers; the size of 85*128 or 128*85. Third a database with 700 images, a total of 7 categories, each category still contains 100 images. Figure 1 shows a sample image database used in our experiments.

In all experiments, are extracted from the image color and texture information as visual features. For the color feature, we will color space transition from RGB to HSV, and the three color channels H, S, V quantization

Figure 1 the sample images used in experiments

for 8 copies, 3 copies and 3 copies. Therefore, the dimension of an image color feature is 72. In addition, the texture features, we will use the Pyramid image wavelet transform (PWT) is the three level of decomposition, and obtains 10 sub zone of the mean and variance, and then get the texture a 20 dimensional feature vector. Finally, we will color and texture features are combined into a 92 dimensional vector, the vector as feature vector of the query image.

In order to verify the validity of OCSVM integration algorithm based on improved AdaBoost, we will be integrated with OCSVM method, the traditional AdaBoost based on SVM, compared OCSVM. In this paper, OCSVM integration method based on AdaBoost for TA-OCSVME, we propose a method for MA-OCSVME. In order to make the above methods achieve better, we use a grid search method to find the optimal parameter values. For SVM, we give the tradeoff parameter C value is 100, the Gauss kernel function K (x, y) = width parameter value of $\exp\{-\gamma \| X - y \|^2\}$ is 2, the optimal results. For the three other methods, the parameter V from {0.2, 0.3, 0.4, 0.5, 0.6} value, while the width parameter from {5, 6, 8, 16, 20, 22} value.

In addition, we will be related to the retrieval process feedback mechanism, to improve the retrieval performance of the system. For every piece of the query image, we have seven feedback, and Euclidean distance as similarity measure.

For each search result, we according to similarity from big to small order image, and returns to the user. The user then image mark similar positive example, but mark not similar negative cases. Because the number of positive cases of database are often less than the negative examples, so we will always positive image marking the last back to the training set, as a training set, in order to improve the performance of classifiers. At the same time, we according to the labeling results rearrange the positive and negative images, thus, probability is occurred in the next to search results will become large, and the probability of negative cases is very small. More to the point, new retrieval results in positive examples often display in front of the. Figure 2 shows a set of cars as a query image and no feedback to the initial retrieval results.

Figure 2 no search results feedback cars

Through experiment summarizes four kinds of methods of average for 50 query image accuracy and feedback frequency. Pay attention to the average of four methods with the increase of the number of accurate feedback have been increasing. Moreover, compared with the other three methods, MA-OCSVME has better retrieval performance. In order to save space, this paper compares the results did not give second and third databases.

Ⅳ Conclusion

In order to improve the performance of image retrieval, we propose OCSVM integration method based on improved AdaBoost used in relevance feedback. Compared with the traditional OCSVM method based on integration of AdaBoost, two SVM, OCSVM comparison, we found that OCSVM integration method based on improved AdaBoost achieves optimal performance.

From the experimental results we can see, the retrieval results, the number of positive cases of image is much less than the number of negative examples of image. The integrated improvement of AdaBoost based on OCSVM can handle this situation of success. In addition,

although the OCSVM can be regarded as a strong classifier, integrated our method achieved the best classification performance, while the traditional OCSVM integration method based on AdaBoost is not satisfactory. In a word, this method effectively improves the retrieval accuracy.

REFERENCES

[1] Y. Rui, T. S. Huang and S. Mehrotra. Content-based image retrieval with relevance feedback in MARS. In International conference on image processing. Santa Barbara, CA, USA, October, 1997, 2: 815-818.

[2] S. Tong and E. Chang. Support vector machine active learning for image retrieval. Proceedings of the Ninth ACM International Conference on Multimedia, Ottawa, 2001, 107-118.

[3] J. Cheng and K. Q. Wang. Active learning for image retrieval with Co-SVM. Pattern Recognition, 2006, 40: 330-334.

[4] D. C. Tao, X. Tang, X. L. Li, et al. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28: 1088-1099.

[5] Y. Chen, X. S. Zhou and T. S. Huang. One-class SVM for learning in image retrieval. Proceedings of the IEEE International Conference on Image Processing, Thessaloniki, Greece, 2001.

[6] S. D. MacArthur, C. E. Brodley and C. R. Shyu. Relevance feedback decision trees in content-based image retrieval. Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries, USA, 2000, 68-72.

[7] J. F. Nunes, P. M. Moreira and J.M.R.S. Tavares. Shape based image retrieval and classification. 2010 5th Iberian Conference on Information Systems and Technologies, 2010, 1-6.

[8] V.S.V.S. Murthy, E. Vamsidhar, J.N.V.R. Swarup Kumar, et al. Content based image retrieval using hierarchical and k-means clustering techniques. International Journal of Engineering Science and Technology, 2010, 2(3): 209-212.

[9] C. D. Ferreira, J. A. Santos, M. A. Goncalves, et al. Relevance feedback based on genetic programming for image retrieval. Pattern Recognition Letters, 2010.

[10] X. S. Zhou and T. S. Huang. Comparing discriminating transformations and SVM for learning during multimedia retrieval. Proceedings of the 9th ACM International Conference on Multimedia. Ottawa, Canada, 2001, 137-146.