International Journal of Hybrid Information Technology
Vol.7, No.5 (2014), pp.341-350
http://dx.doi.org/10.14257/ijhit.2014.7.5.31

Indoor and Outdoor Scene Classification Method Based on Fourier Transform

Zhijie Zhao, Haitao Wang, Xuesong Jin, Huadong Sun and Qian Wu

School of Computer and Information Technology,
Harbin University of Commerce
Harbin, Heilongjiang, China
zhijiezzz@aliyun.com, martinkeith@163.com

Abstract

In this paper, a method for indoor and outdoor scene classification is proposed based on the spectrum information of Fourier transform. Firstly, the image is divided into five partitions. The color, texture, and spectral information of each partition are extracted as feature vector. Each partition has a separate image feature vectors, so the image can be represented by this set of features. For the feature extraction stage, frequency information can be used to classify indoor and outdoor scene. Compared with other classification methods, classification accuracy of this method increased from 91% to 92%, experimental results show that the proposed method is effective.

Keywords: *Image Classification; Indoor-Outdoor Scene; Fourier Spectrum; SVM*

1. Introduction

Indoor and outdoor scene classification play an important role in the image processing, Such as content-based image recovery [1, 2], digital library [3] and digital photography [4]. The main difficulty is that there are many similar content in the indoor and outdoor scene, such as plants or the building structure. There are some effective classification methods which using different classifiers, features, training methods, test data and semantic knowledge. Texture and color spatial features are frequently used. Single feature can not be used for indoor and outdoor scenes classification sufficiently. In order to produce better classification results, two or more features are often combined as feature vector [6-8].

In some methods, images are divided into blocks, and then these blocks are independently analyzed to reduce the computational complexity. Szummer and Picard [6], Serrano, *etc.*, [7, 9] divided the image into two equal-sized portions with named two pass system. In the first pass, each part is labeled as indoor or outdoor using color and texture separately. The second pass combines the results of the first pass to make an overall indoor/outdoor classification on the image. Similarly, several sub-blocks are connected together as a feature for classification [10, 11].

Color and texture feature are frequently used for indoor and outdoor scene classification problems [12]. Navid Serrano *et al.*, [12] obtained color feature through histogram. Besides, other features are effective for classification such as Wavelet texture feature, the statistical feature from the color histogram. These texture features are shown to perform better than previously used texture features. Wavelet texture feature have been obtained through two-wave layers of wavelet decomposition.

ISSN: 1738-9968 IJHIT

In addition to selecting the appropriate features, choosing a good classifier is also important. At present, some classification methods such as KNN classifier [1, 6, 16, 17], Bayesian analysis method [10, 11], support vector machines [3, 7, 9] and forward neural network [13, 14] are used. The appropriate parameters for these classifiers can be calculated from features.

It has been proven that KNN classifier is inefficient and cannot get an accurate K from the validation set [7]. Bayesian approaches are very difficult in general pattern recognition and neural networks are usually hard to optimize for generalization [18]. The classifier performance often depends on the amount of training data and optimization strategy. Since, it is hard to compare the classification result from different image data and optimize learning.

The method proposed by Jim *et al.*, [20] taking into account that most of the middle parts of the image have little effect on the classification results, for example human faces and body etc. Therefore, this method is proved to have better accuracy. However, it considers only the color and texture edge features. These are all features of the spatial domain. But the frequency domain features for scene classification also has an important role. Based on the characteristics of the original, using the Fourier transform spectrum of the image as the classification features, we tried to further improve the accuracy of classification.

2. Edge and Color Orientation Histogram Algorithm

2.1. Image Segmentation

In most of images, the center of image area is the target object (such as faces), obviously, these objects cannot be used to distinguish types of indoor and outdoor scenes. The edge region compared to the central can get more information of the type of scene. Therefore, the image is divided into five parts, as shown in Figure 1. BLK1, BLK2, BLK5 is defined as edge area of the image, BLK3, BLK4 is the central area. For each small block, edge orientation histogram (EOH) and color orientation histogram (COH) should be calculated.

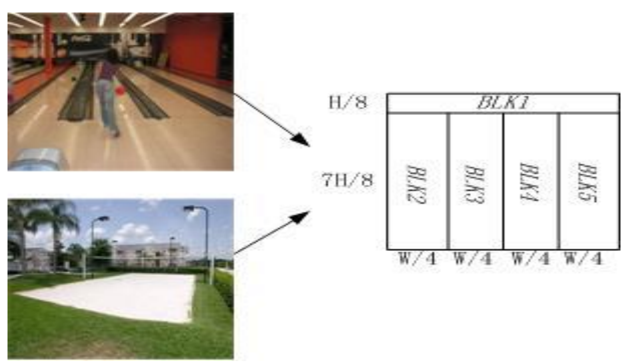


Figure 1. The Initial Segmentation of Image

2.2. Edge Orientation Histogram (EOH)

For each (i = 1...5), using the following formula, the amplitude and amplitude angle of the texture edge pixels are calculated.

$$A(x, y) = \sqrt{P_x(x, y)^2 + P_y(x, y)^2} \tag{1}$$

$$\theta(x, y) = \tan^{-1} \frac{P_y(x, y)}{P_x(x, y)} \tag{2}$$

Where A (x, y) is the amplitude, $\theta(x, y)$ is amplitude angle, $P_x(x, y)$ and $P_y(x, y)$ are the horizontal and vertical direction vector of pixel (x, y). Vector direction of the pixel is quantized range from 0 ° to 180 °. 180 degrees is divided into 8 equal portions. According to the amplitude angle of texture edge pixels, the edge pixels are divided into eight regions. The sum of amplitudes of all pixels within each region is calculated. The following formula:

$$E_{i,m} = \sum_{\substack{(x,y)\in BLK_i \\ \theta(x,y)\in m}} A(x, y), 1 \le i \le 5, 1 \le m \le 8 \quad (3)$$

Here m (x, y) and θ (x, y) denote the edge magnitude and quantized orientation at the pixel position (x, y), respectively.

2.3. Color Orientation Histogram (COH)

For color orientation histogram is different from with EOH. The H and V components of HSV color space are used. According to the range of V component, *i.e.*, 0 ° -360 °, the V component is divided into 8 equal portions, the sum value of H within each part is calculated. The following formula:

$$C_{i,n} = \sum_{\substack{(x,y)\in BLK_i \\ \theta(x,y)\in m}} S(x, y), 1 \le i \le 5, 1 \le m \le 8 \quad (4)$$

Here, s (x, y) and h (x, y) represent the saturation and hue at pixel (x, y). The ECOH descriptor is finally defined by combining the EOH and COH descriptor for each block. The feature vector is generated by multiplying weights for each block to ECOH descriptor. The feature vector is shown in the following equation.

$$F = \omega_1 F_1^{ECOH}, \omega_2 F_2^{ECOH}, \cdots \omega_5 F_5^{ECOH}$$

Where,

$$F_i^{ECOH} = F_{i1}^{E} F_{i1}^{C}, F_{i2}^{E} F_{i2}^{C}, \cdots, F_{i8}^{E} F_{i8}^{C}, 1 \le i \le 5 \quad (5)$$

3. Based on the Fourier Transform Spectrum Extracting Feature

Most of the outdoor scenes contain relatively flat texture features, such as the sky. For indoor scene objects, its complexity, variety and relatively less space, it will contain more edges and linear information.

This characteristic is reflected to the frequency domain. For frequency spectrum, the outdoor scenes relatively change strongly. Based on this characteristic, the frequency information combined with ECOH, trying to get more accurate results.

3.1. Fourier Transform of the Image

The discrete Fourier transform (FT) of an image is defined as:

$$I(f) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} i(x) h(x) e^{-j2\pi\langle f,x\rangle} \quad (6)$$

The amplitude spectrum is defined $A(f) = |I(f)|$. The amplitude spectrum reveals the dominant orientations and textural patterns in the image (Figure 2).

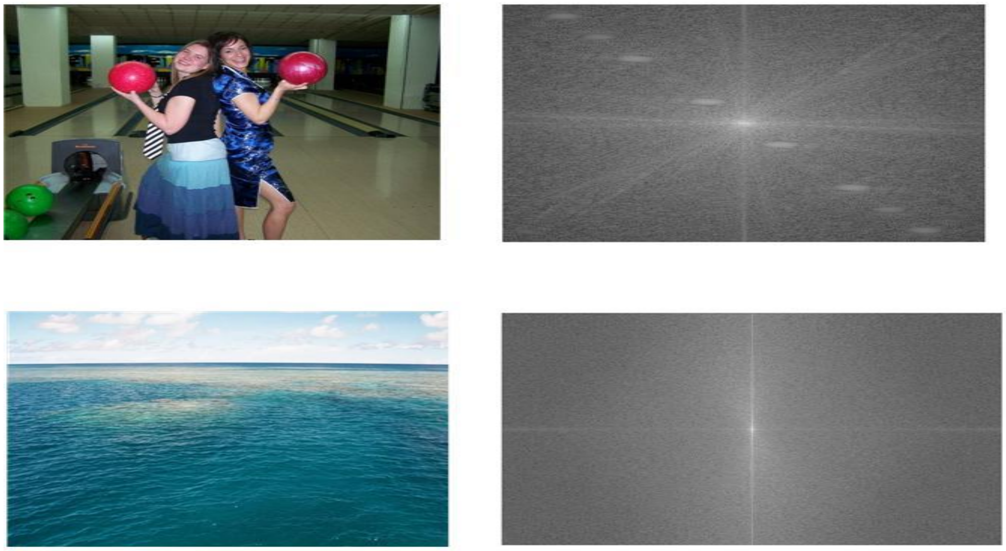


Figure 2. The Left Column is the Two Example Images used in this Paper. Right Column is the Magnitude Spectrum of the Fourier Transform

3.2. Signal Smoothness and Fourier Transform Relationship

The paper will discuss the relationship between signal smoothness and Fourier transform mode attenuation. Considering a set of samples:

$$(1)\ f(x) = e^{-x^2/2},\ x \in (-\infty, +\infty)$$

$$(2)\ f(x) = e^{-|x|},\ x \in (-\infty, +\infty)$$

$$(3)\ f(x) = \begin{cases} e^{-|x|}, & x \in (0, +\infty) \\ 0, & x \in (-\infty, 0] \end{cases}$$

Three functions have different degrees of smoothness. a) Smooth everywhere (Traditionally, smooth means infinitely differentiable, but sometimes does not discuss in detail), b) smooth everywhere but the point of x=0, still continuous at x = 0; c) smooth everywhere but the point of x=0, and no-continuous at x = 0. Intuitively, the example a) is the smoothness of the best, followed by b), and finally c). If the Fourier spectral analysis is used to determine the smoothness of the signal, the Fourier transform of the three functions, namely:

$$(1)\ F(\omega) = (2\pi)^{1/2} e^{-\omega^2/2}$$

$$(2)\ F(\xi) = \frac{1}{1+\omega^2}$$

$$(3)\ F(\psi) = \frac{1}{1+\omega^2} + j\frac{\omega}{1+\omega^2}$$

Typically, the most convenient way is to directly compare the Fourier-mode (amplitude spectrum). The first two functions are the transform themselves. The last one is $\left|F(\omega)\right| = (\frac{1}{1+\omega^2})^{\frac{1}{2}}$. It is easy to determine that the stronger singularity signal is the slower mode of Fourier decay. Figure 3 shows the amplitude spectrum of these functions.

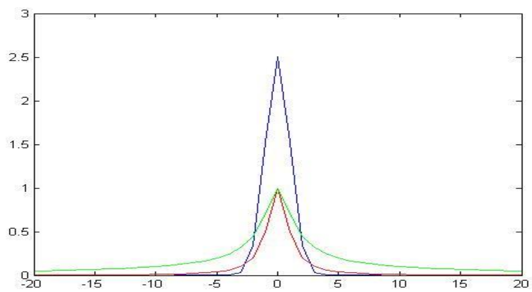


Figure 3. Attenuation of the Signal Amplitude Spectrum of Different Smoothness

The above analyses indicate that the more smoother the signal is, the faster the attenuation of it in the frequency domain is. Figure 4 shows the real part of the Fourier transform of the real signal:

$$\int_{-\infty}^{+\infty} f(x)\cos x\omega\, dx$$

The imaginary part:

$$\int_{-\infty}^{+\infty} f(x)\sin x\omega\, dx$$

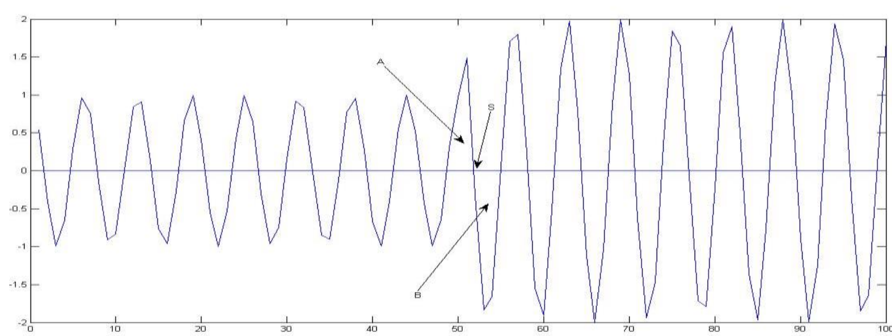


Figure 4. Discontinuous Point Makes the Fourier Integral of the Signal Appear Slow Decay

Assuming that it occurs discontinuity at the point of x = s (The point is similar to the image edge). In the Figure 4, two areas which named A and B are not equal just because of discontinuity, so they cannot be canceled out each other. Usually local singularity place (such as edges) is called the local high-frequency regions. It can be seen from this explanation, in fact, the attenuation of the Fourier transform modulus can judge the signal appears intermittently or other non-smooth phenomena. For the sky part of the outdoor scenes, the signal is smooth. According to the above analysis of the known, Fourier Transform spectrum can be used as a distinguishing feature of indoor and outdoor scenes.

3.3. The Difference of Fourier Spectral Curve of Indoor and Outdoor Image

Indoor and outdoor scenes have different degree of smoothing. It shows that Fourier spectrum as a classification feature is very effective. To evaluate this relationship, first of all, we introduce kurtosis of the curve.

Kurtosis is also known as the kurtosis coefficients. Kurtosis formula is as follows:

$$k = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^4}{(N-1)S^4} \tag{7}$$

The unit of Kurtosis is the BK. $\overline{Y}$ is average value of n times measured samples. S is the sample standard deviation. Kurtosis of the normal distribution is 3.bk<3 is called the distribution has insufficient kurtosis, bk>3 is called distribution has excess kurtosis. Briefly, kurtosis is the description of the steepness of distribution patterns. If the Kurtosis is 3BK, it is the same as the normal distribution; Kurtosis more than 3 indicates distribution is sharper than the normal distribution, while Kurtosis less than 3 means smoother distribution.

After transforming and translating the images in the frequency domain, spectrum curve can be achieved. They cannot represent in the form of specific functions. It is difficult to analysis. From the perspective of the analysis of differences spectral curves, discrete is continuous sampling. For convenience, the above mentioned several functions and Fourier Transform are still considered. Variable range is taken from - 100 to 100, and seeking kurtosis respectively, kurtosis values are: 79.5801, 86.3241 and 2.1441. According to the nature of kurtosis, kurtosis values of normal function is 3, so kurtosis values of smooth function is larger ,and non-smooth function is smaller.

So based on the method ECOH, new scenes classification features is constructed through adding the Fourier spectrum information, the feature vector is shown in the following equation:

$$F = F_1^{ECOH}, F_1^{O}, F_2^{ECOH}, F_2^{O} \cdots F_5^{ECOH}, F_5^{O}$$

Where, $F_i^{O} = F_{i1}^{O}, F_{i2}^{O} \cdots F_{i32}^{O}\ 1 \le i \le 5$

3.4. Algorithm Flow

1. The picture is segmented into five blocks, and calculating texture edges for each block, according to the orientation of the pixel texture, the amplitude of the pixel texture edge is divided into eight bins, then determine the sum amplitude of pixels in each bin , and as a dimension of the feature space, the orientation of texture is $0^\circ \sim 180^\circ$,since angles between 180° and 360° can be considered the same as angles in $0^\circ \sim 180^\circ$, i.e. mod 180.
2. The image is converted from the RGB color space to HSV color space, as the range of the H component is $0^\circ \sim 360^\circ$, While the S component represents color saturation, Therefore, it is similar to the first step (process of seeking EOH), 360 degrees is divided into eight equal portions, based on the H component of the pixel values,

different pixels are divided into 8 bins, determine the sum of S component in each bin, and as a dimension of the feature. 80-dimensional vector can be obtained through above two calculations, the form of feature space is introduced earlier, not repeat them here.

3. For each block (i=1~5), apply Fourier transform, then shift for making the part of low frequency in center.
4. Above all, after the translation transform, the spectral distribution of indoor or outdoor images is different. Therefore, each block is segmented again. By the way of statistical histogram, Spectrum regular pattern of indoor and outdoor scenes are figured out. several methods are tried, better results can be achieved when each block segmented into 32 small blocks (Figure 5). Among them, the first block is segmented into 2 rows, each row has 16 blocks. The other four large blocks is segmented into 16 rows, and each row has 2 blocks. As the size of images is not consistent, after dividing the large block equally, the part which less than one equal block is abandoned as it has little impact on the classification results.
5. After segmentation, all pixels within each of the small block are calculated the amplitude then sum as a dimension of the feature vector.

The spectral characteristics of the image is extracted through the above five steps, finally, each image can possess 240-dimensional feature vectors through adding spectral characteristics to the original method of ECOH.

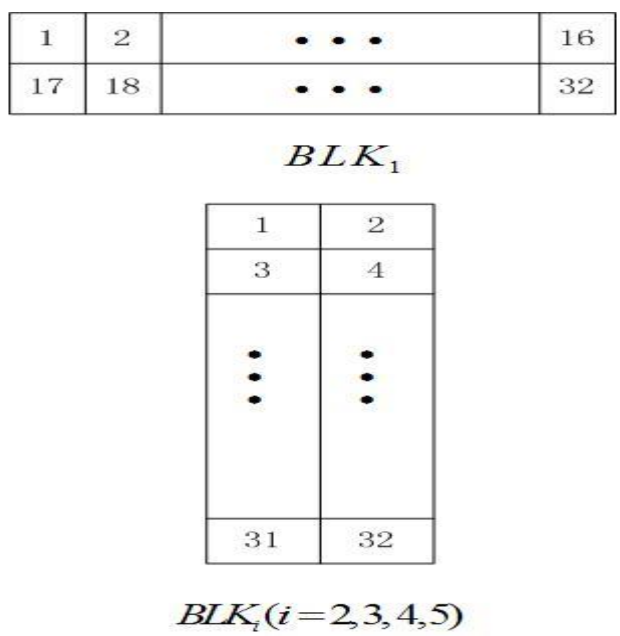


Figure 5. The Method of Statistics the Distribution of Frequency Spectrum

4. Experimental Results

In order to demonstrate the effective of the proposed method in this article, different types of indoor and outdoor images are selected (such as trees, buildings, beaches, airports, supermarkets, bedroom, and so on). The experimental image database comes from Corel image database [19] and Department of Computer Science and Engineering University of Washington [20]. There were 2,638 training sample images, the number of indoor images is

1494 , outdoor images is 1144 ,and there were a total of 2139 test sample images, the number of indoor images is 1583 , outdoor images is 556. Matlab and libsvm Toolbox is used [21] to complete experiments. In order to get more accurate results, parameter optimization is performed before establishing the SVM Model. C and gamma are two important parameters in SVM model. C is the coefficient of punishment, which means tolerance level of error. The higher the value c is the more intolerable errors. C is too large or too small, generalization ability become poor. Gamma is a parameter that comes with RBF function after that RBF function was selected as the kernel. Distribution is determined implicitly after that data is mapped to the new feature space by gamma. The bigger of Gamma value result in the less support vector, the smaller of Gamma value result in the more support vector. The number of support vector influences the speed of training and prediction. So it's important for classification results to select C and gamma reasonably. First, feature space after training is equally segmented into 5 parts. Optimal value of C and gamma were found by the way of the cross-validation, namely,4 parts of five parts are used as training samples, the left part is used as test samples. Different values of C and gamma will lead to different predicted classification accuracy. According to the classification accuracy, it's identified eventually that c is 4, gamma is 1, and predicted classification accuracy through using different values of C and gamma are given in the following table:

Table 1. ECOH Method Optimization

$\log_2(C)$	$\log_2(gamma)$	***predicted accuracy***
16	-30	74.2%
16	4	65.75%
24	-16	85.55%
2	0	90.55%

Table 2. Method of the Paper Optimization

$\log_2(C)$	$\log_2(gamma)$	***predicted accuracy***
16	-30	74.45%
16	4	64.55%
24	-16	84.6%
2	0	90.55%

And Figure 6 shows all of the prediction accuracy with different c and gamma.

Optimization was performed in Python. Experimental verification was completed by ECOH and method proposed in this article, the results obtained are as follows Table 3:

Table 3. Accuracy Compared

method	***accuracy***
ECOH	91%
This paper	92%

It is shown by the results, the method for indoor and outdoor scenes classification which using frequency domain spatial information is feasible and effective.

5. Conclusions

In this paper, a method using Fourier Transform spectrum information was proposed for indoor-outdoor image classification. Images were first segmented into five parts. The spectrum information was combined with the ECOH method, and support vector machines were used for classification. Accuracy was increased 1% compared with ECOH method. High accuracy of method proposed in this paper can be applied to scene classification for indoor and outdoor images. Other transform spectrum features (such as the wavelet transform) will be taken into account in future work, and improving classification accuracy.

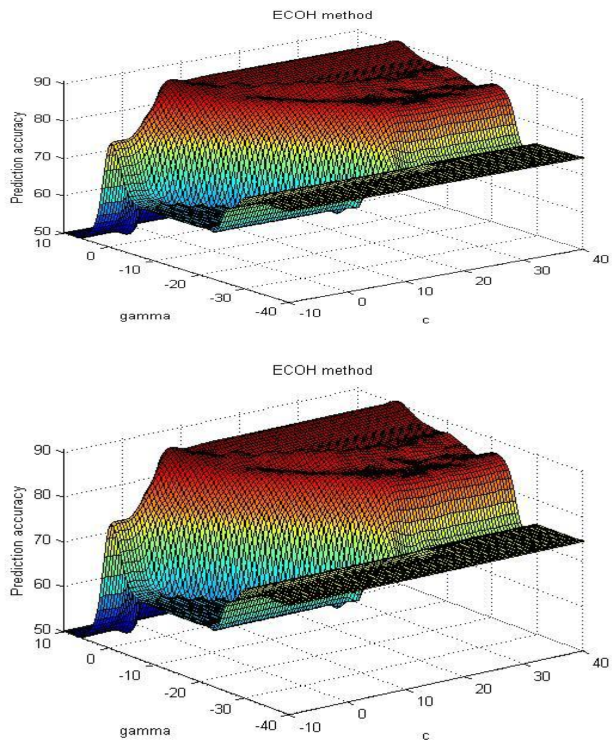


Figure 6. Final Figure of Two Method Optimization

Acknowledgements

This work was supported by Technology Foundation for Selected Overseas Chinese Scholar, the technological innovation foundation of Harbin (2012RFQXG090), and Natural Science Foundation of Heilongjiang Province (F201114 and F201245).

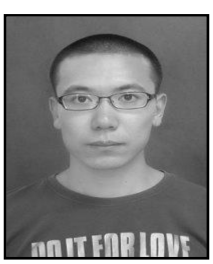
References

[1] G. Eason, B. Noble and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions”, Phil. Trans. Roy. Soc. London, vol. A247, **(1955)** April, pp. 529-551.
[2] G. Nathalie, L. B. Herve, H. Jeanny and G. D. Anne, “Towards the Introduction of Human Perception in a Natural Scene Classification System”, IEEE Conference NNSP, Switzerland, **(2002)** September 4-6.
[3] G. Qiu, X. Feng and J. Fang, “Compressing Histogram Representations for Automatic Colour Photo Characterization”, Pattern Recognition, vol. 11, **(2004)**, pp. 2177-2193.
[4] S. Srinivasan and L. Kanal, “Qualitative Landmark Recognition using Visual Cues”, Pattern Recognition Lett., vol. 18, **(1997)**, pp. 1405-1414.

[5] Y. Su and F. Jurie, "Improving Image Classification Using Semantic Attributes", International Journal of Computer Vision, vol. 100, no. 1, **(2012)**, pp. 59-77.
[6] M. Szummer and R. Picard, "Indoor-outdoor image classification", IEEE International Workshop on Content-Based Access of Image and Video Databases, ICCV'98, **(1998)**.
[7] N. Serrano, A. Savakis and J. Luo, "Improved scene classification using efficient low-level features and semantic cues", Pattern Recognition, **(2004)**, in press.
[8] J. Luo, A. Savakis, "Indoor vs. outdoor classification of consumer photographs using low-level and semantic features", Proceedings of the International Conference of Image Processing, Thessaloniki, Greece, **(2001)**.
[9] N. Serrano, A. Savakis and J. Luo, "A computationally efficient approach to indoor/outdoor scene classification", IEEE International Conference on Pattern Recognition, **(2003)**.
[10] A. Vailaya, M. Figueiredo, A. K. Jain and H. J. Zhang, "Content based hierarchical classification of vacation images, IEEE Multimedia Conference, Florence, Italy, **(1999)**.
[11] A. Vailaya, M. Figueiredo, A. K. Jain and H. J. Zhang, "Image classification for content-based indexing", IEEE Transactions on Image Processing, vol. 10, **(2001)**, pp. 117-130.
[12] S. Bianco, G. Ciocca, C. Cusano and R. Schettini, "Improving Color Constancy Using Indoor-Outdoor Image Classification", IEEE Transactions on Image Processing, vol. 17, no. 12, **(2008)** December.
[13] M. Veta, T. Kartalov and Z. Ivanovski, "Content-based Indoor/Outdoor Video Classification System for a Mobile Platform", World Academy of Science, Engineering and Technology, **(2009)**.
[14] J. Mtimet and H. Amiri, "Image Classification using Statistical Learning Methods", Journal of Software Engineering and Applications, **(2012)**, pp. 200-203.
[15] N. Serrano, A. Savakis and J. Luo, "A computationally efficient approach to indoor-outdoor scene classification", 16th International Conference on Pattern Recognition, vol. 4, **(2002)**, pp. 146-149.
[16] P. Fitzpatrick, "Indoor/outdoor scene classification project", http://www.mit.edu/ people/paulfitz/pub/indoor-outdoor.pdf.
[17] Q. Iqbal and J. K. Aggarwal, "Retrieval by classification of images containing large manmade objects using perceptual grouping", Pattern Recognition, Workshop on Applications of Computer Vision, vol. 35, **(2002)**, pp. 1462-1479.
[18] A. Savakis and J. Luo, "Indoor vs. Outdoor classification of consumer photographs", Proceedings of the IEEE International Conference on Image Processing, ICIP, Thessaloniki, Greece, **(2001)**.
[19] L. Antanas, M. Hoffmann, P. Frasconi and T. Tuytelaars, "A Relational Kernel-based Approach to Scene Classification", IEEE Workshop on Applications of Computer Vision (WACV), **(2013)**.
[20] W. Kim, J. Park and C. Kim, "A Novel Method for Efficient Indoor–Outdoor Image Classification", Signal Processing Systems, vol. 61, no. 3, **(2010)**, pp. 1-8.
[21] http://wang.ist.psu.edu/docs/related/.
[22] http://www.cs.washington.edu/research/imagedatabase.
[23] http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

Authors

Zhijie Zhao, Doctor, Professor, IEEE Member, His research fields are image processing and intelligence information processing.



Haitao Wang, Harbin University of Commerce post-graduation, His research fields are computer networks and communications.

