Visual Features Fusion for Scene Images Classification

GAO Hua Member, IAENG and ZHAO Chun-xia, ZHANG Hao-feng

***Abstract*—Scene images classification is one of the most important methods in robot navigation in unstructured environment. In this paper, a scene images classification method based on visual feature fusion is proposed. First, the scene image is divided into fixed-size blocks. Second, computes the DCT coefficients of each blocks and runs a wavelet-like decomposition, means and variances are extracted in each component of the decomposed sub-blanks. Third, color moments are extracted and fused with DCT texture-based features to describe each block. Finally, a Gaussian Mixture Model (GMM) is employed for training and classification.**

***Index Terms*—Scene Classification, Robot Navigation, Gaussian Mixture Model (GMM), Discrete Cosine Transform (DCT), Color Moments**

I. INTRODUCTION

With the development of computer vision, vision navigation has become one of the most important parts of autonomous navigation in nature outdoor scenes for unmanned ground vehicles (UGVs) and intelligent robots. Autonomous robots capture visual information of scenes and train a classifier, predict the taversability of real scene images with the trained classifier. The promising classification results provide the basis for obstacles avoiding and path planning.

Roberto Manduchi et al. [1]-[3] analyzed the affects of illumination changes on terrain classification and compensated the illumination deficiencies, then proposed a color-based classification method on several typical terrain, such as soil/rock, green vegetation, and dry vegetation. Jansen et al. [4] assumed that images with large similarities in environment related properties such as illumination, materials and geometry also have similar pixel distributions in a color space. They distinguished training image sets that share environmental states and separately modeled the terrain type colors for each image set to reduce the influence of changing environmental conditions. It was concluded that color is a very effective feature for terrain typing. The soil and dry grass classes, however, are incorrectly classified due to inherent color ambiguities existing between those classes. They assert that color can be used in conjunction with other features in order to enhance delineation of these classes.

Recently, classification based on texture has been addressed by many researchers. Pietikäine et al. [5] proposed a texture classification method multi-local binary pattern (LBP), Castano et al. [6] used Gabor filters bank to extract texture and proposed an obstacles detection method, and SoundraPandian et al [7] proposed a traversability classification method using Gabor filters bank. Gi-Yeul Sung et al. [8] presented a terrain classification method based upon color and texture features, they extracted texture features with discrete wavelet transform coefficients mean and energy values. Furthermore, they adopted spatial coordinates where a terrain is located in the image as additional features. Shi X et al. [9], [10], Permuter et al. [11] proposed their own traversability detection method based on fusion of color feature and texture feature respectively.

In this paper, a scene classification method is proposed in this paper. First, texture features are extracted in gray channel. Then, color moments are computed in all three color channels of RGB color space. The parameters of GMM are estimated by training these labeled samples. Finally, new scene images are classified by the trained GMM model.

The rest of the paper is organized as follows: Gaussian Mixture Model (GMM) method is given in section 2. Section 3 shows the color feature and DCT-based texture extraction methods, defines the distance of fusion feature vectors. Section 4 presents our results and finally section 5 deals with conclusions.

II. GAUSSIAN MIXTURE MODEL

Gaussian Mixture Model (GMM) is an extended version of single Gaussian probability density function, can approximate arbitrary shaped densities. Given a data set $\{\boldsymbol{x}_i\}_{i=1,2,\ldots,n}$, $\boldsymbol{x}_i \in \boldsymbol{R}^d$, if the distribution of these data is approximately ellipsoidal, the probability density function of these data can be written as follows:

$$g(x;\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)^d|\Sigma|}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \quad (1)$$

Manuscript received Junuayr 8, 2012; this work was supported by China National Nature Science Funds 90820306 and 61101197.

GAO Hua received the B.S. and M.S. degrees from Shandong University at Weihai and Nanjing University of Science and Technology, in 2005 and 2007, respectively. Currently, he is pursuing the Ph.D degree in the College of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, China. His current research interests include pattern recognition, intelligent robot, machine learning, image processing. (phone: +86 18951738200; e-mail: ghua2009@gmail.com).

ZHAO Chun-xia received the B.S., M.S. and Ph.D. degrees from Harbin Institute of Technology, in 1985, 1988 and 1998, respectively. Since 1998, she has been with Nanjing University of Science and Technology, Nanjing, China, where she is currently a Professor in the College of Computer Science and Technology. Her current research interests include pattern recognition, image processing, artificial intelligence and intelligent robot. (e-mail: zhaocx@mail.njust.edu.cn).

ZHANG Hao-feng received the B.S. and Ph.D. degrees from Nanjing University of Science and Technology, in 2003 and 2007, respectively. Since 2007, he has been with Nanjing University of Science and Technology, Nanjing, China, where he is currently an Associate Professor in the College of Computer Science and Technology. His current research interests include machine learning, pattern recognition, and intelligent robotics. (e-mail: haofeng_njust@163.com).

Where μ is the center of density function and Σ is the covariance matrix of density function. Using Maximum Likelihood Estimation (MLE) method, parameters μ and Σ can be worked out.

The distribution function of arbitrary d-dimensional data can be written as weighted average of several Gaussian density functions:

$$p(x)=\sum_i \alpha_i g(x;\mu_i,\Sigma_i)$$
$$\sum_i \alpha_i = 1 \qquad (2)$$

This is so-called Gaussian Mixture Modal. Utilizes the same MLE method can work out all its parameters μ_i and Σ_i.

III. Visual Feature Extraction and Fusion

A. DCT features

A discrete cosine transform (DCT) expresses a sequence of finitely many data points in terms of a sum of cosine functions oscillating at different frequencies. Given a 2D data set $\{\boldsymbol{A}_{m,n}\}_{m=1,\dots,M,n=1,\dots,N}$, the DCT coefficients are calculated by the following formulas:

$$B_{p,q}=\alpha_p\alpha_q\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}A_{m,n}\cos\frac{\pi(2m+1)p}{M}\cos\frac{\pi(2n+1)q}{M} \qquad (3)$$

Where, $\alpha_p=\begin{cases}1/\sqrt{M} & p=0\\ \sqrt{2/M} & 1\le p\le M-1\end{cases}$, $\alpha_q=\begin{cases}1/\sqrt{N} & q=0\\ \sqrt{2/N} & 1\le q\le N-1\end{cases}$ are the normalized factors.

In this paper, we discarded small high-frequency components to smooth the data. Then run a wavelet-like decomposition, as shown in Fig.1.

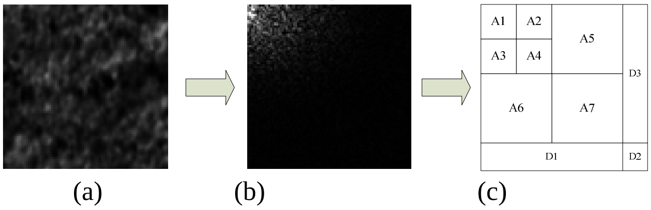


Fig.1 Image of DCT coefficient smoothness and decomposition. (a)source image; (b)DCT coefficients; (c) DCT coefficient smoothness and decomposition

As illustrated in Fig.1, small high-frequency components (region D1, D2, D3) are discarded, they are not included in the calculation of features. The remained components are decomposed into sub-bands, as A1~A7 in Fig.1(c), just like wavelet decomposition.

After the DCT coefficient smoothness and wavelet-like decomposition, the means and variances of each sub-band are calculated to combine the texture feature vector. The discrete cosine transform coefficients means and variances of the i^{th} sub-band are calculated using the following formulas:

$$m_i=\frac{1}{|A_i|}\sum_{(p,q)\in A_i}|B_{p,q}| \qquad (4)$$

$$s_i=\sqrt{\frac{1}{|A_i|}\sum_{(p,q)\in A_i}\left(|B_{p,q}|-m_i\right)^2} \qquad (5)$$

Where $|A_i|$ is the number of DCT coefficients of i^{th} sub-band, $|B_{p,q}|$ is the absolute value of DCT coefficients $B_{p,q}$. It's a 2-level decomposition in Fig.1(c), while different level decomposition generates different number of features. If the level num is denoted by M, the number of sub-bands is ($3M$+1). Combine DCT coefficients means and variances of all sub-bands, the texture vector of M-level decomposition is $\boldsymbol{F}_{texture}=[m_1,s_1,m_2,s_2,\dots,m_{3M+1},s_{3M+1}]$.

B. Color moments

For a color image, the i-$^{\text{th}}$ order color moment [12] is defined by

$$F_i=\begin{cases}\dfrac{1}{A}\sum_j x_j & i=1\\ \left(\dfrac{1}{A}\sum_j |x_j-F_1|^i\right)^{1/i} & i\ge 2\end{cases} \qquad (6)$$

Where, A is the area of feature extraction window, x_j is the $j^{\text{-th}}$ pixel. In practical, several low-order color moments can describe the color feature enough. This paper selects the first 3 order color moments of each color channels of RGB color model to describe the color feature of each blocks, $\boldsymbol{F}_{col}=[F_{1,R},F_{2,R},F_{3,R},F_{1,G},F_{2,G},F_{3,G},F_{1,B},F_{2,B},F_{3,B}]$.

C. Feature fusion

In this paper, we combine color-moments and DCT-based texture features to describe the blocks. The final form of block feature is $\boldsymbol{x}=[\boldsymbol{F}_{col},\boldsymbol{F}_{texture}]$.

Mahalanobis distance is a distance measure based on correlations between variables, gauges similarity of an unknown sample set to a known one. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant. Formally, the Mahalanobis distance of a multivariate vector from a group of values with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is defined as:

$$d_M(x,y)=\sqrt{(x-y)^T\Sigma^{-1}(x-y)} \qquad (7)$$

According to formula (7), the texture distance and color distance can be written as follows:

$$d_{text}(x,y)=\sqrt{[x_{text}-y_{text}]^T\Sigma_{text}^{-1}[x_{text}-y_{text}]}$$
$$d_{col}(x,y)=\sqrt{[x_{col}-y_{col}]^T\Sigma_{col}^{-1}[x_{col}-y_{col}]} \qquad (8)$$

In this paper, we defined the normalized Mahalanobis distance as follows:

$$\begin{aligned}d_{NM}(x,y)&=\sqrt{\lambda d_{text}^2(x,y)/d_{text}+(1-\lambda)d_{col}^2(x,y)/d_{col}}\\&=\sqrt{(x-y)^T\begin{bmatrix}\lambda d_{text}^{-1}\Sigma_{text}^{-1} & 0\\ 0 & d_{col}^{-1}(1-\lambda)\Sigma_{col}^{-1}\end{bmatrix}(x-y)}\end{aligned} \qquad (9)$$

Where, d_{text} and d_{col} are the dimension of texture feature vector and color feature vector respectively. The magnitude of distance relays on the vector's dimension largely, the high-dimensional texture features always have more contribution than low-dimensional color features for dissimilarity. In this paper, we normalized the both distance measures by dividing by their dimensions d_{text} and d_{col}. $\lambda\in[0,1]$ is the balance parameter of the color features and texture features for dissimilarity contribution. Then the formula (1) can be written as:

$$\begin{aligned}&g(x;\mu,\Sigma)\\&=\gamma\exp\left(-(x-y)^T\begin{bmatrix}\lambda d_{texture}^{-1}\Sigma_{texture}^{-1} & 0\\ 0 & d_{color}^{-1}(1-\lambda)\Sigma_{color}^{-1}\end{bmatrix}(x-y)\right)\end{aligned} \qquad (10)$$

Where,

$$\gamma = \frac{1}{\sqrt{\lambda(1-\lambda)(2\pi)^{d_{col}+d_{text}}\left|\Sigma_{text}\right|\left|\Sigma_{col}\right|/\left(d_{text}d_{col}\right)}}\,.$$

IV. Experiments

A. Experiment in VisTex database

We chose 9 textured color images from the Vistex database, which size is 512 × 512, as shown in Fig.2. The feature extraction was done on blocks of size 18, with overlap of 2 pixels, we had 1024 blocks in each image. After discrete cosine transform, remained 16×16 low-frequency coefficient matrix of 18×18 DCT coefficient matrix for texture feature extraction in each block. In each image, the first 256 blocks were used for training, while the remaining 768 blocks were used for testing. We selected the balance parameter λ=0.5 and decomposition level M=3.

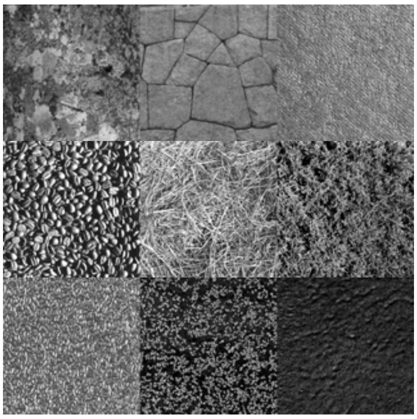


Fig.2 Source images from Vistex database

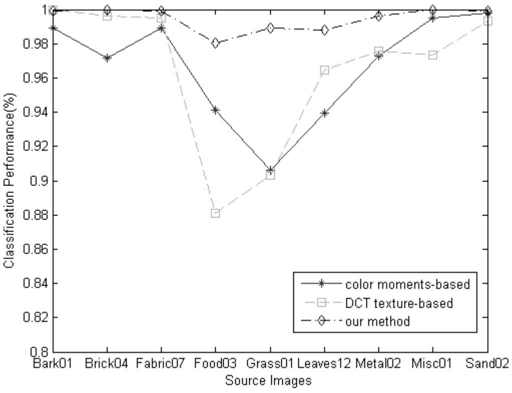


Fig.3 Classification rate in different source images

Fig.3 shows the classification results of color moments-based method, DCT texture-based method and our fusion method. All three methods perform well in images Bark01, Brick04, Fabric07, Metal02, Misc01, Snd02, which are showed in 1st and 3rd row of Fig.2. But color moments-based method and DCT texture-based method perform poor in images of 2nd row in Fig.2, especially in image Food03 and Grass01. Our method performs in those three images not as well as in other six images, about 1.3% lower, but better than methods based on single feature, about 10% higher in image Food03 and Grass01.

B. Experiment in our scene images

We captured a sequence of real off-road images, and cut 16 sub-images for training which size was 81×81. 4 typical terrains were considered, which were sky, road, grass and tree, 4 images in each scene, as shown in Fig.4. First row of Fig.4 shows the 4 sample images of sky, and second row shows road samples, third row shows grass samples, fourth row shows tree samples. The feature extraction was done on blocks of size 9, DCT-based texture features were calculated in 8×8 sized low-frequency components and decomposition level M was 2.

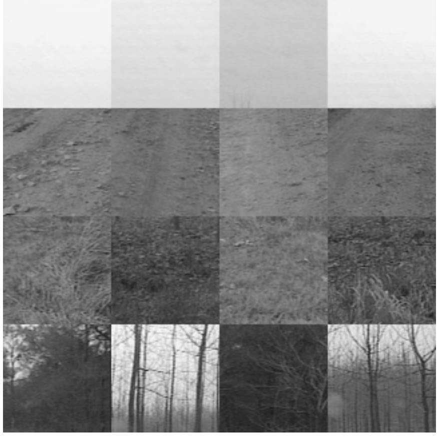
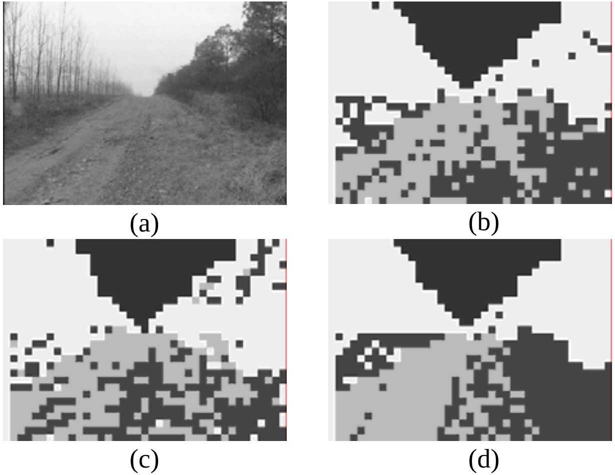


Fig.4 Training samples of experiment in our scene images

Fig.5 shows the classification results of our scene images, (a) is the test image, (b) and (c) are the results of color-based method and DCT texture-based method respectively, and (d) is the result of our fusion method. All those three methods perform well in sky detection and tree detection, there are little sky is misclassified in color-based method and DCT texture-based method. But there is much misclassification in two other scenes. Color-based method and DCT texture-based method can only distinguish those clear roads. Our fusion method performs not as well in grass scenes and road scenes as in sky scenes and tree scenes, can't detect the road perfectly, but detects a traversable path and distinguishes more regions than the other two method.



■sky ▨road ■grass □tree

Fig.5 (a)source image; (b)result of color-based method;
(c)result of DCT texture-based method; (d)result of our method

V. Conclusion

This paper proposed a scene classification method based on fusion of color feature and texture feature. Texture features are extracted by calculated the means and variances of low-frequency DCT coefficient sub-banks. Color moments are calculated from three color channels of RGB color model to combine color features. Normalized Mahalanobis distance measure is utilized to compute the dissimilarity of feature vectors and balance the magnitude of the contribution of texture features and color features for vector dissimilarity. And then a balance factor is exploited to decide the ratio of color features and texture features.

We compared our method with color moments-based method and DCT feature-based method. The results show that our method performs better than those two methods in VisTex database and our real scene images. Even in complex environments in which there is much misclassification in our method, our method performs better than those two methods.

REFERENCES

[1] Roberto Manduchi, “Obstacle detection and terrain classification for autonomous off-Road navigation,” Autonomous Robots, 2005, vol, 18, no, 1, pp. 81-102.

[2] Roberto Manduchi, “Learning Outdoor Color Classification”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28, no, 11, pp. 1713-1723.

[3] Roberto Manduchi, “Obstacle detection and terrain classification for autonomous off-road navigation,” Autonomous Robots, 2005, 18, pp. 81-102.

[4] Paul Jansen, Wannes van der Mark, Johan C. van den Heuvel, Frans C.A. Groen, “Colour based Off-Road Environment and Terrain type Classification,” In Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems, pp. 216-221. Vienna, Austria (2005)

[5] Pietikäinen M, Nurmela T, Mäenpää T, et al., “View-based recognition of real-world textures,” Pattern Recognition, 2004, 37, pp. 313-323.

[6] Castano R, Manduchi R, Fox J. “Classification experiments on real-world textures,” In Proceedings of Workshop on Empirical Evaluation in Computer Vision, Kauai, HI, 2001, pp. 1-20

[7] SoundraPandian K K, Mathur P., “Traversability assessment of terrain for autonomous robot navigation,” In Proceedings of the International MultiConference of Engineers and Computer Scientists 2010, Hong Kong: IMECS 2010, 2010, pp. 1286-1289

[8] Gi-Yeul Sung, Dong-Min Kwak, Joon Lyou, “Neural Network Based Terrain Classification UsingWavelet Features,” Intelligent Robot System (2010), vol, 59, pp. 269-281.

[9] Shi X, Manduchi R. “On the Bayes fusion of visual features,” Image and Vision Computing archive, 2007, vol, 25, no, 11, pp. 1748-1758.

[10] Shi X, Manduchi R. “A study on Bayes feature fusion for image classification,” Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition Workshop, 2003, pp.1-9.

[11] Haim Permuter, Joseph Francos, Ian Jermyn, “A study of Gaussian mixture models of colour and texture features for image classification and segmentation,” Pattern Recongnition, 2006, vol,39,no,4,pp.695-706.

[12] Shih Jau-Ling, Chen Ling-Hwei, “Color Image Retrieval Based on Primitives of Color Moments,” Recent Advances in Visual Information Systems (2002), pp. 19-27.