Study on Fire Early Warning System Based on Sparse Coding Visual Bag-of-words Model

Jiajun Wang*1, Yahui Wang*2

*1 Energy Efficiency Valuation Department, China Machinery Technical Development Foundation, Beijing, China，100053
E-mail: wjiajun1990@sina.com
*2 School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China
E-mail: yahui-wang@vip.sina.com

Abstract: Based on the early design and implementation of wireless multimedia sensor networks (WMSNs), the mobile network platform for fire early warning was established for verifying the problem of using video monitoring to directly identify fire to avoid the misjudgment of traditional fire smoke and temperature detector. Aiming at the contradiction between time-delay and real-time of massive data transmission in WMSNs, an algorithm for image information big data processing was studied and applied to fire warning. Simulation and experimental results have shown that this strategy was effective. The research focus on the image preprocessing at the acquisition end of wireless network sensors, and an algorithm based on sparse coding visual word package model multi-attribute feature fusion was proposed. First, the fire target was detected, and the compressed image information was transmitted according to the detection results and decision-making requirements. This new processing strategy of fire image based on WMSNs has greatly reduced the transmission burden of WMSNs (Wireless Multimedia Sensor Networks), improved the data transmission rate and enhanced the real-time performance of information. The experimental verification results have also shown that the model was effective for target recognition. The classification decision learning algorithm of support vector machine (SVM) was trained with positive and negative samples, and the experimental verifications of the decision function were conducted, and the results were accurate.

Keywords: Fire Early Warning System, Sparse Coding, Bag-of-words (BOW) Model

1. INTRODUCTION

According to complexity of the firefighting system, an intelligent fire warning system is constructed by using internet of things technology and image and video processing technology [1]. This early warning system can give full play to the advantages of mobile internet of things, such as multi-node, flexible network topology, and fault tolerance of routing, so as to greatly improve the accurate early warning of specific situations of fire, fundamentally changing the limitations of the traditional fire early warning system, such as smoke sense can only identify smoke and smolder [2]. Early fire warning information is intuitive, the fire point location is accurate, and the development situation of the on-site fire can be dynamically monitored at any time, thus improving the rationality and scientific nature of the mobilization, allocation and use of emergency resources, so as to realize more economical and efficient fire protection, rather than the traditional "at all costs".

Using image recognition method, fire fighters can intuitively and accurately judge the fire scene, make fire control and rescue plans timely and correctly, improve the efficiency of fire control work, and enhance the ability of emergency management [3]. Wireless multimedia sensor network (WMSNs) has been widely applied in various fields [4]. However, when applied to the transmission of fire image information, it is necessary to solve the contradiction between the transmission time-delay and real-time of a large amount of data generated by Multimedia information. Therefore, in the image preprocessing of wireless network sensor information acquisition, an algorithm based on sparse coding visual word packet model multi-attribute feature fusion was proposed to extract and compress image features, which greatly reduced the information transmission burden of WMSNs. At present, the dynamic monitoring of the fire situation on site has been realized, which greatly improves the accuracy and reliability of the fire early warning. Simulation and experimental results have shown that this strategy is effective.

2. WIRELESS MULTIMEDIA SENSOR NETWORK FIRE DETECTION STRATEGY

In order to realize convenient, quick and reliable application goal in WMSNs at the scene of the fire detection in wireless multimedia network fire detection based on mobile terminal system, a new wireless multimedia sensor network fire data acquisition and transmission control strategy was proposed based on the theory of digital image processing, aim at the problem of low efficiency of large quantities of data, which greatly reduce the burden of the WMSNs transmission, and simplify the amount of data transmission, improve the efficiency of data transmission[5].

The procedure of this strategy is shown in Fig. 1.

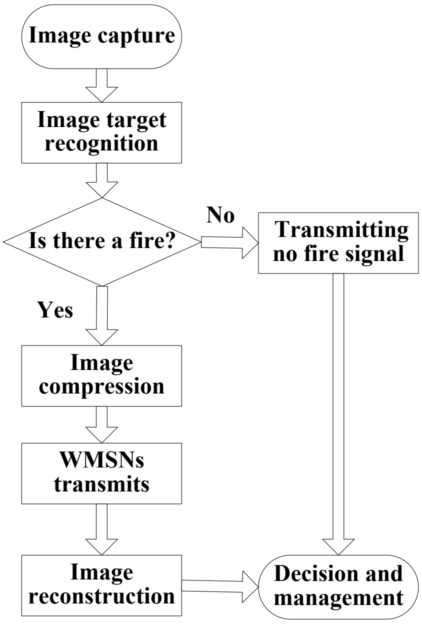


Fig.1. Fire detection strategy for wireless Multi-media sensor network

This data transmission strategy focuses on the terminal sensor acquisition module of the perception layer. For the collected field scene image information, it was preprocessed in the terminal acquisition module to judge the nature of the information and carry out sub-transmission, so as to reduce the amount of effective data transmission and improve the efficiency of information transmission [2]. The specific processing steps were as following: the images of the surrounding environment were collected by image detector, in which the fire characteristic information was detected by the image recognition algorithm. The situation of non-fire or fire identified by the signal identifier was determined respectively, and the identification signal was sent to the management system. In case of fire, the collected field fire images were further compressed and transmitted, and the images were reconstructed after arriving at the remote management system to obtain the specific and detailed information of the fire scene. If there is no fire, no further transmission of image information was required. Remote users can also, on demand, issue commands to transmit images to obtain situation information about the current environment.

In the above process, a target recognition algorithm based on sparse coding visual word package model and multi-attribute feature fusion is designed to detect the fire information on image and determine whether there is a fire. In other words, the visual codebook was constructed by sparse coding firstly, then the sliding window was fully described by the late fusion algorithm of the word package model, and finally the target existence was determined by the classification of sliding window region by support vector machine (SVM).

3. TARGET DETECTION ALGORITHM BASED ON SPARSE CODING MULTI-ATTRIBUTE VISUAL BOW

3.1 Design of Fire Target Recognition Algorithm

Target recognition algorithm based on sparse coding visual word package model multi-attribute feature fusion was adopted. The specific flow chart is shown in Fig. 2.

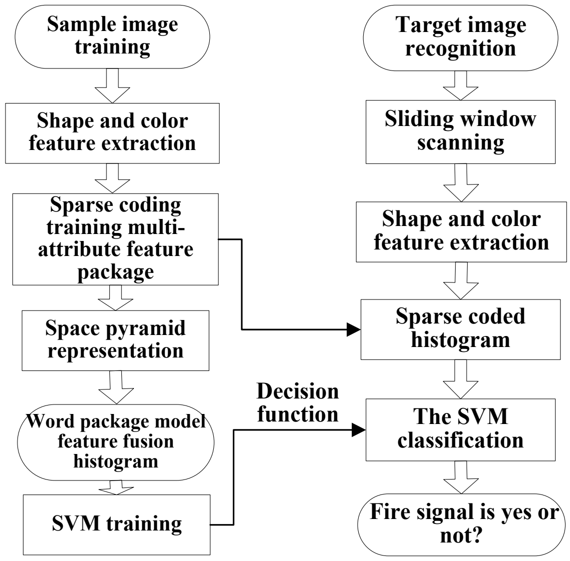


Fig.2. Flow chart of target image detection

Identification process was mainly divided into two stages [6]:

(1) The training stage: There are both positive and negative samples in training, firstly basing on gradient-based optimization algorithm to extract the local characteristics of the training sample images, and then building shape and color features visual word package respectively by the fusion of color and shape features [7] strategy to improve the performance of target detection. At the same time, the sparse coding principle was adopted to identify effective data and reduce data transfer. The spatial pyramid vector descriptor based on shape and color was calculated according to the maximum statistical principle, and the histogram description of the training image was constructed by using the late feature fusion algorithm. Finally, the classification decision function was obtained by training SVM.

(2) The recognition stage: The target image extracts each child window's partial color and shape feature by sliding windows to compare with training process' that. According to the sparse coding principle and multi-characteristics data fusion algorithm, the child

window's Multiple Attribute Feature Fusion Pyramid histogram descriptor is collected, and also according to the SVM decision function, whether the detection image's child window is the target area is determined.

3.2 Linear Support Vector Machine Multiple Classification Algorithm

Support vector machine (SVM) is a classifier used commonly, which takes training error as optimization constraint condition and minimizes structural risk as learning criterion. SVM has very strong generalization ability, which can be extended from two categories to multiple categories. It is widely used in image recognition, information compression and other fields. In this paper, bag-of-words model (BOW) histogram was adopted to classify model scenes based on word packages and SVM was used to identify fire information.

The decision function of SVM linear classifier was defined as equation (1):

$$f(z)=\sum_{i=1}^{n}\alpha_i\kappa(z,z_i)+b \tag{1}$$

Where, $\kappa(z,z)$ is Mercer kernel function , corresponding to the inner product of a certain transformation space, the linear classification after a nonlinear transformation can be realized, and can be selected according to different circumstances. In this paper, a multi-level linear SVM was used to classify the spatial pyramid descriptions of sliding subwindows, and the training sample was assumed to be $\{(z_i,y_i)\}_{i=1}^{n}$, $y_i\in Y=\{1,\cdots,L\}$. First, SVM must learn L linear functions $\{w_c^T z\,|\,c\in Y\}$ to obtain the SVM classification hyperplane, and then calculate the classification label of data features z according to the following formula equation (2).

$$y=\max_{c\in\xi} w_c^T\cdot z \tag{2}$$

The theoretical basis of SVM is nonlinear mapping, the target is the optimal hyperplane, and the result is the support vector. SVM is to transform the input space into a high-dimensional space, in which the optimal classification surface is obtained. The feature space is divided into two parts by the hyperplane. Therefore, training SVM requires large storage space and high computational cost, which greatly limits the efficiency of classification. However, SVM's decision function is determined by a small number of support vectors, avoiding the "dimensional disaster", and at the same time has good robustness, suitable for solving small sample, nonlinear and high-dimensional pattern recognition and other problems, and can also be applied to function fitting and other fields.

If $y_i=c$, then $y_i^c=1$, otherwise $y_i^c=-1$, where $l(w_c,y_i^c,\,z_i)$ denotes the hinge loss function. The standard loss function is not differentiable everywhere and has a great impact on the gradient-based optimization algorithm. Quadratic differentiable loss function can be used to realize efficient gradient algorithm training SVM, which is defined as follows equation (3) and (4).

$$\min_{w_c}\left\{J(w_c)=\|w_c\|^2+C\sum_{i=1}^{n}l\left(w_c;\,y_i^c,z_i\right)\right\} \tag{3}$$

$$l\left(w_c\,;y_i^c,\,z_i\right)=\left[\max\left(0,w_c^T z,y_i^c-1\right)\right]^2 \tag{4}$$

Where, C represents the punishment variable.

3.3 Construction of Visual BOW based on Sparse Coding

Sparse coding is a mathematical representation method for linear decomposition of multidimensional data. K-means clustering is to determine the initial clustering center for each cluster of the sample set, and then the samples is distributed to the center vector closest to them, so as to achieve the minimum value of the objective function [8].

For construction of Visual bag-of-words model based on sparse coding[9], suppose $X=\left[\mathrm{X}_1,\cdots,\ \mathrm{X}_\mathrm{M}\right]^\mathrm{T}\in\mathrm{R}^{\mathrm{M}\times\mathrm{D}}$, represents the set of feature descriptors extracted in the image area, M and D represents the feature dimension and the number of features respectively. Vector quantization optimization is carried out by using k-means clustering method, then equation (5) as

$$\min_{\mathrm{V}}\sum_{m=1}^{M}\min_{k=1\cdots K}\left\|X_m-\mathrm{V}_k\right\|^2 \tag{5}$$

Where, $\mathrm{V}=\left[\mathrm{V}_1,\cdots,\ \mathrm{V}_\mathrm{K}\right]\in\mathrm{R}^{\mathrm{M}\times K}$ was codebook set composed of K cluster centers. If equation (1) was converted into the least-square fitting form with constraints, then the coefficient u_m and corresponding constraint conditions are added to obtain equation (6):

$$\min_{\mathrm{U,V}}\sum_{m=1}^{M}\left\|\mathrm{X}_m-\mathrm{u}_m\mathrm{V}\right\|^2 s.t.\ Card\left(\mathrm{u}_m\right)=1, \\ |\,\mathrm{u}_m|=1,\mathrm{u}_m\geq 0,\forall m \tag{6}$$

Where, $\mathrm{U}=\left[\mathrm{u}_1,\cdots,\ \mathrm{u}_\mathrm{m}\right]^\mathrm{T}$ represents the quantization matrix with the local feature descriptor, and the cluster center of the vector X_m can be determined by the unique non-zero element value in u_m .

Using the sparse coding optimization algorithm, a sparse regular term was added in equation (6) to weaken the constraint conditions $|\mathrm{u}_m|=1$, then equation (7) as:

$$\min_{U,V}\sum_{m=1}^{M}\|X_m-u_mV\|^2+\lambda|u_m| \qquad (7)$$
$$subject\ to\ |v_k|\le 1,\forall k=1,2,\cdots,K$$

Where, λ is weakening coefficient. In the traditional BOW model, various local features of the image were represented by visual vocabularies, and the joint distribution vector of multiple visual vocabularies was used to represent the local features of the image, which reduces the quantization error.

3.4 Sparse Coded BOW Space Pyramid Representation

The spatial pyramid structure model is shown in Fig.3.

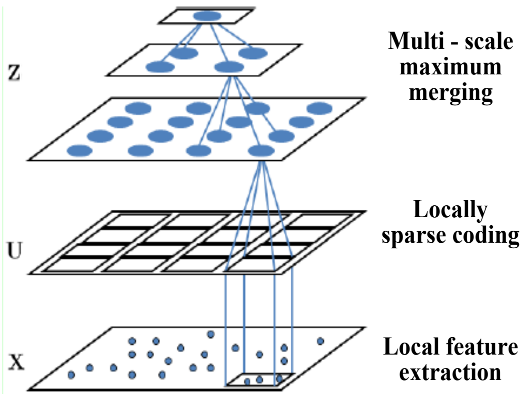


Fig.3. Sparse coding space pyramid model diagram

Where, X is a set of feature descriptors extracted from the image region, V for a visual word package code set constructed in advance by the optimization algorithm, and U for a mapping matrix from local features of the image region to all visual vocabularies V_m in the codebook. The vector descriptor of the image region is represented by Z, then the joint function F was constructed, so that equation (8) was as:

$$Z=F(U)\quad Z=[z_1,z_2,\cdots z_K]^T\in R^{K\times 1} \qquad (8)$$

Z is a K-dimensional descriptor. Common joint functions F are square root of mean square (Sqrt), absolute mean value (Abs) and maximum value (Max). In this paper, the maximum joint function was adopted, and j elements in the vector descriptor Z was for equation (9).

$$z_j=\max\{|u_{1j}|,|u_{2j}|,\cdots,|u_{Mj}|\} \qquad (9)$$

Where, u_{ij} is the element in i row and j column of the matrix U, and M is the number of local features in the image region.

3.5 Multi-attribute Feature Fusion Algorithm based on BOW

There are currently two image local feature fusion algorithms based on the BOW model both Early Fusion and Late Fusion [10].

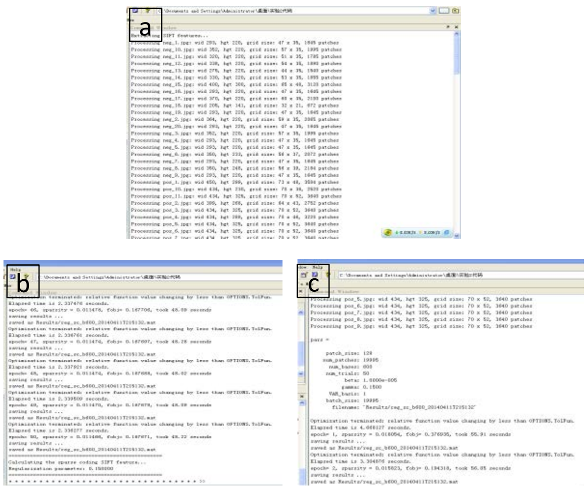
The late fusion method in image local feature fusion algorithms based on the BOW model is adopted. Assume that the image $I_i(i=1,2,\cdots N)$ feature descriptor was $f_{ij}(j=1,2,\cdots,M^i)$, M^i is the number of local features in i's the image region. and the visual vocabulary in the constructed multi-attribute feature visual word package was w_i^k, $i=1,2,\cdots V^k$, $k\in\{s,c,sc\}$, where s and c respectively represented the shape and color visual vocabulary, and SC represented locally fused shape-color visual vocabulary. The content of the image can be described by frequency distribution histogram of visual vocabulary, so the calculation formula of frequency distribution of each visual vocabulary is in equation (10):

$$n(w^k\,|\,I^i)=\sum_{j=1}^{M^i}\delta(w_{ij}^k,w^k)$$
$$\delta(x,y)\begin{cases}1 & for\ x==y\\ 0 & for\ x\neq y\end{cases} \qquad (10)$$

The late fusion method $f_{ij}=\{w_{ij}^s,w_{ij}^c\}$ in image local feature fusion algorithms based on the BOW model is adopted. The visual lexical frequency distribution of shape $n(w^s\,|\,I^i)$ and color $n(w^c\,|\,I^i)$ are calculated respectively, and then the frequency distribution histograms of the two attributes are cascaded by weighting, so the accuracy of image recognition is improved.

4. EXPERIMENTAL SIMULATION AND ANALYSIS

Two groups of samples, positive and negative, were used to train SVM, the initial value setting and training process was as follows figure 4:



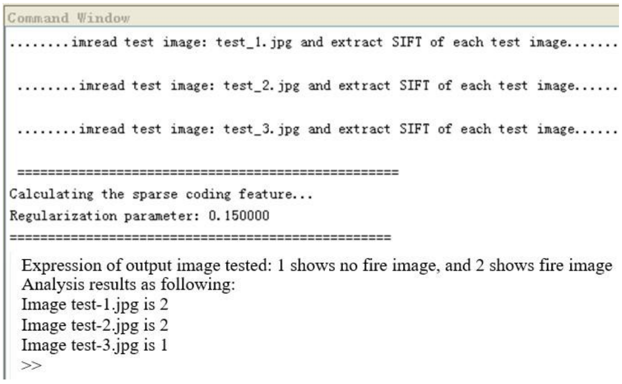
a) Reading training samples b) Setting SVM initial parameters c) Ending SVM training

Fig.4. Displaying on SVM training processes

a b c

a test-1.jpg b test-2.jpg c test-3.jpg
Fig.5. Living example No.1 on fire detection images recognition verification [11]

The Recognition verification result of fire detection images about Example No. 1 was obtained in figure 6.

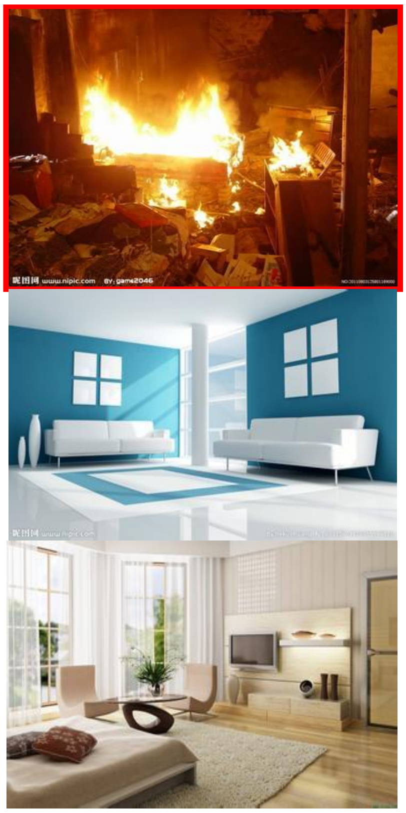


Command Window
........imread test image: test_1.jpg and extract SIFT of each test image.......
........imread test image: test_2.jpg and extract SIFT of each test image......
........imread test image: test_3.jpg and extract SIFT of each test image......
==
Calculating the sparse coding feature...
Regularization parameter: 0.150000
==
Expression of output image tested: 1 shows no fire image, and 2 shows fire image
Analysis results as following:
Image test-1.jpg is 2
Image test-2.jpg is 2
Image test-3.jpg is 1
>>

Fig.6. Recognition verification result of fire detection images about Example No. 1

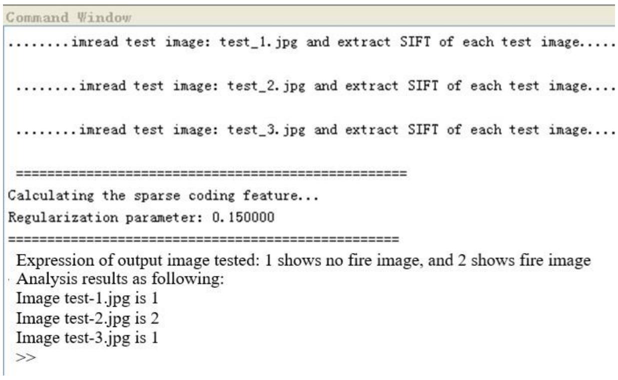
(2) For the second time, the detection images were in figure7.

a b c



a test-1.jpg b test-2.jpg c test-3.jpg
Fig.7. Living example No.2 on fire detection images recognition verification [11]

The Recognition verification result of fire detection images about Example No. 2 was obtained in figure 8.



Command Window
........imread test image: test_1.jpg and extract SIFT of each test image.....
........imread test image: test_2.jpg and extract SIFT of each test image....
........imread test image: test_3.jpg and extract SIFT of each test image....
==
Calculating the sparse coding feature...
Regularization parameter: 0.150000
==
Expression of output image tested: 1 shows no fire image, and 2 shows fire image
Analysis results as following:
Image test-1.jpg is 1
Image test-2.jpg is 2
Image test-3.jpg is 1
>>

Fig.8. Recognition verification result of fire detection images about Example No. 2

5. CONCLUSION

i. The intelligent fire warning system was preliminarily constructed by using internet of things technology and image and video processing technology. The early warning system can give full play to more nodes, flexible network topology, routing, fault tolerance and a series of advantages of the mobile internet of things, so as to achieve early warning information intuitive, fire point positioning accurately, fire development at any time to dynamically monitor the situation on the spot, greatly raising the accuracy of the specific situation for fire warning, and to improve the emergency resource mobilization, configure, and use of rationality and scientific nature.

ii. In view of the multimedia internet transfer of huge amount of data and image video signal transmission delay and distortion of signal, the scene images in situ were preprocessed using visual word package based on sparse coding model of image information processing technology. According to the demand of the fire control work, the selective extraction and screening of the fire scene image characteristics were conducted to ensure that only effective information transmission, greatly accelerate the data transfer rate, and enhance the real time information.

iii. In order to demonstrate the effectiveness of sparsely coded multi-attribute feature package model for target recognition, the experiments were conducted for several times. After that, the classification decision learning algorithm SVM was trained with both positive and negative samples, so as to conduct experimental verification of the decision function, and the results are accurate.

REFERENCES

[1] Li Rui, Wang Yahui, Wang Jianling, “Research on SQLCE based on Zigbee Wireless Sensor Network System,” Intelligent Building Electrical Technology, 2013, 7(2): pp99-102.

[2] Enyan Sun, Xuanjing Shen, Haipeng Chen, Chuanyun Wang, “Distributed Image Compression and Transmission Scheme in Wireless Multimedia Sensor Networks,” TELKOMNIKA: Indonesian Journal of Electrical Engineering, 2013,12 (1): pp661-668.

[3] Shi Wenwu, Yang Jun, “Application of wireless Sensor Network in Greenhouse Environmental Monitoring,” Computer and Modernization, 2012, 2012(3):pp125-127.

[4] Liu Fengyu, “Research on cross-layer Model and QoS Routing of Wireless Multimedia Sensor Networks,” Changsha: The Master Degree Dissertation of Hunan University, 2009.

[5] Ji Lin, “Fire detection based on image Feature Extraction,” Dalian: The Master Degree Dissertation of Dalian University of Technology, 2013.

[6] Navneet Dalal, Bill Triggs, “Histograms of Oriented Gradients for Human Detection,” Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

[7] Shahbaz Khan F, Anwer R M, van de Weijer J, “Color Attributes for Object Detection,” Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012: pp3306-3313.

[8] Maytham Safar, Hasan Al-Hamadi, Dariush Ebra-himi, “PECA: Power Efficient Clustering Algorithm for Wireless Sensor Networks,” International Jo-urnal of Information Technology and Web Engin-eering, 2011, 6(1): pp49-58.

[9] Yang J, Yu K, Gong Y, “Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification,” 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Miami, Florida, 2009: pp1794-1801.

[10] Shahbaz Khan F, Van De Weijer J, Vanrell M, “Top-down Color Attention for Object Ecognition,” Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 2009: pp979-986.

[11] Wang Jianling, “Wireless Sensor Network System Integration” Beijing: The Master Degree Dissertation of Beijing University of Civil Engineering and Architecture, 2014