*Available online www.jocpr.com*

**Journal of Chemical and Pharmaceutical Research, 2014, 6(7):1333-1340**

**Research Article**

**ISSN : 0975-7384**
**CODEN(USA) : JCPRC5**

# Obstacle detection and road segmentation by 3-D reconstruction based on monocular vision

**Yu Hong[1], Wang Zhengyou[2] and Hong Ruxia[1]**

[1]*Department of Information Science, Nanchang Teachers' College, Nanchang, China*
[2]*Department of Electronic and Information Engineering, Shijiazhuang Tiedao University, Shijiazhuang, China*

## ABSTRACT

*Using monocular systems in computer vision has many advantages including low cost, high mobility, etc. In this paper we reconstruct 3-D spatial model using single still image so as to fulfill the task of obstacle detection and road segmentation. In our solution, we represent 3-D spatial model as two important parts: spatial layout and object depth. Firstly, we present an algorithm to annotate objects in the images such as road, buildings, then we utilize a state-of-the-art depth estimation algorithm to obtain depth map from the image, and an SVM is used to combine spatial layout and object depth so as to get obstacle detection result. Finally, the experimental results show that this method is simple and effective which has high practical value.*

**Key words:** Obstacle detection; Road segmentation; Monocular; 3-D reconstruction; Depth.

## INTRODUCTION

As one important subject in the machine vision task, obstacle detection and road recognition technology is one of current research hot spots. Now the solutions extensively used in the related research achievements include ultrasonic, laser distance measurement or binocular parallax for 3-D Reconstruction [1-3]. However, most of these methods involve valuable devices and larger volume and are susceptible to environmental condition, so they are not very suitable for medium- and small-sized robots. The obstacle detection and road recognition based on monocular system features low cost and high efficiency and is gradually emphasized by foreign and domestic scholars. In [4], it segments the image to get the obstacle area by using LVQ NN classifier. He etc [5] proposes a small target threshold selection algorithm to segment images and get the obstacle and road area simultaneously. In [6], it matches and tracks the image sequence feature points for obstacle detection. In [7], it proposes a method to detect the target obstacles by computing the shift vector of the road feature points. The above method collects image data and processes them by using single camera. Its algorithm only uses the partial image sequence data. Compared to single static image, these algorithms should be further improved in accuracy and reliability. To realize true monocular vision, this paper aims to detect obstacles and segment valid roads in the image by using single static image.

To achieve the above targets, we reestablish 3-D spatial model from single image. In essence, the called 3-D spatial model reflects a series of important space properties in the image scenario, e.g. distribution of different objects such as buildings, vehicles and roads, and their relative positions from observers. Given an image, this paper describes 3-D spatial model of the image scenario by using space layout and depth. first, this paper proposes an image scenario labeling method based on the space layout information to get the road and potential obstacle area, next we predict the depth of objects in the images by using the depth prediction algorithms of objects, and finally we combine the space a SVM with the space layout information and depth information to get the obstacle area in the image. The experiment research indicates that our model is highly adaptive and accurate and is suitable for many scenarios indoor and outdoor.

## 1. DEPTH PREDICTION ALGORITHM OF OBJECTS

As everybody knows, it is not enough to only simply consider the features such as local color and texture of images in prediction of the depth of objects. E.g. when observing a blue image block, we do not know that it is heaven or blue object, so we cannot determine the depth. Both local image feature and global space relation should be considered. In [8], Saxena proposes a method to predict the depth of objects in the images. This method can get the image features via the following processes. First, segment an image into many small blocks and describe each small block with the absolute depth feature and relative depth feature. The absolute depth feature describes the depth of a specific small block and the relative depth feature describes depth difference of two adjacent small blocks.

The absolute depth feature is acquired as follows: Given an image block *i* in the image $l(x,y)$, compute the 9-D Laws mask texture [9], 2-D color channel and 6-D text gradient to form a 17-D filter $F_n(x,y)$, $n=1,2,\cdots,17$. $E_i(n)=\sum_{(x,y)\in each(i)}|I(x,y)\times F_n(x,y)|^k, k=\{1,2\}$, indicates accumulative absolute value and square sum outputted by the filter, so a 34-D initial feature vector is obtained. Next a multi-scale model is used to survey feature of adjacent blocks under different scales and features of column with this specific block. For detailed description, refer to the figure 1. Finally we get a 646-D ($19\times 34$) feature to describe specific block in the image.

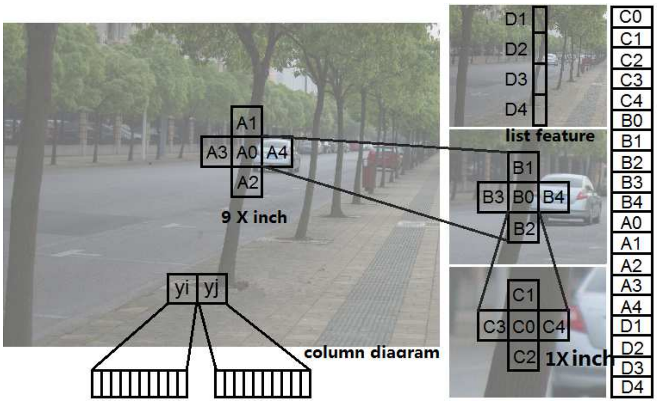

**FIGURE 1: Final absolute and relative feature vector obtained by surveying a specific image block and its adjacent blocks under multi-scale model**

The relative depth feature is computed as follows: for each image block *i*, one 10-bin full histogram is used to describe output of each dimension of the filter $|l(x,y)\times F_n(x,y)|$, so we get a 170-D ($17\times 10$) feature $y_i$ in order to survey how adjacent image blocks are associated. Finally the histogram difference $y_{ij}=y_i-y_j$ of two image blocks is the relative depth feature of the image block *i* and *j*.

After the absolute depth feature of each block and relative depth feature of each adjacent block group in the image are obtained, the local image feature is fused with the space relation via the Markov random field (MRF). Here $s=1,2,3$ indicates different scales of images in the model.

$$d_i(s+1)=\frac{1}{5}\sum_{j\in\{i,N_s(i)\}}d_j(s)$$

$N_s(i)$ is 4 adjacent blocks of the image *i* under the dimension *s*, so MRF model is described as follows:

$$P(d\mid X;\theta,\ \sigma)=\frac{1}{z}\exp(-\sum_{i=1}^{M}\frac{(d_i(1)-x_i^T\theta_r)^2}{2\sigma_{1r}^2}-\sum_{s-1}^{\theta}\sum_{\mathrm{i}=1}^{M}\sum_{j=N_i(t)}\frac{(d_i(s)-d_j(s_j))^2}{2\sigma_{2rs}^2})$$

*M* is the sum of the minimum dimension blocks in the image. $x_i$ is the absolute depth feature of the image block *i*. $\theta,\sigma$ are the model parameters. *Z* is the normalization factor. The conditional probability $P(d\mid X;\theta_r)$ with maximum training set is used to estimate the parameter $\theta_r$. In addition, $\sigma_{2rs}^2=u_{rs}^T\mid y_{ijs}\mid$, $\sigma_{1r}^2=v_r^Tx_i$ are estimated. $u_{rs} > 0$ and $v_r > 0$ are a group of parameters which can fit $\sigma^2_{2rs}$ to $(d_i(s) - d_j(s_j))^2$ and fit $\sigma^2_{1r}$ to $(d_i(r) - \theta_r^{\mathrm{T}} x_i)^2$.

In [8], the system collects 425 true depth diagrams of the scenario by using a customized 3-D laser scanner. 75% images are used to train this MRF model and residual 25% images are used for test. Given a test image, after the image features are extracted, the MAP estimate of the depth of the field of the scenarios is obtained via the maximization equation above. The figure 2 shows depth diagram of two groups of true image depth diagrams and depth diagram predicted by using this model.

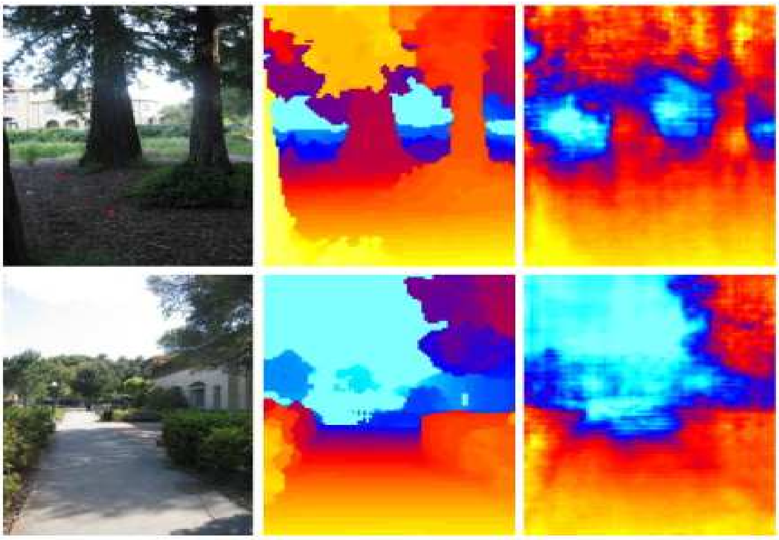

**FIGURE 2: Depth diagram**

*(Left: original diagram, middle: true depth by using a laser scanner, right: depth diagram computed by a model)*

## 3. SPACE LABELING ALGORITHM

### *3.1 Overview*

Some research indicates that the images collected via common mobile carrier have the following features: the sight line is horizontal, the image has 3-D affine features and most pixels can be divided into heaven, ground and middle scene. In addition, the middle scene can be divided into different orientations. To determine object orientation in an area, the long line of this area can be extracted and the crossover point of these lines is computed, namely affine center. The collected images has 3-D affine feature, so the artificial objects such as building and furniture in the scene have an affine center. The affine disappearing point of the left object is on the left side of the image, the affine disappearing point of the right object is on the right side of the image, and the affine disappearing point of the object orienting to the observer is inside or above and under the image. In addition, partial natural objects such as hill and wood have irregular shapes and have no affine features, next label a pair of input images as the heaven, ground and middle scene. The middle scene is further divided into left, right, observer-oriented and non-orientation scene. We finally expect to get the ground area. Generally the obstacles orient to the observer or have no specific orientation, so the space layout labeling is favorable to the final obstacle detection and road segmentation. The figure 3 shows our labeling structure of the image.

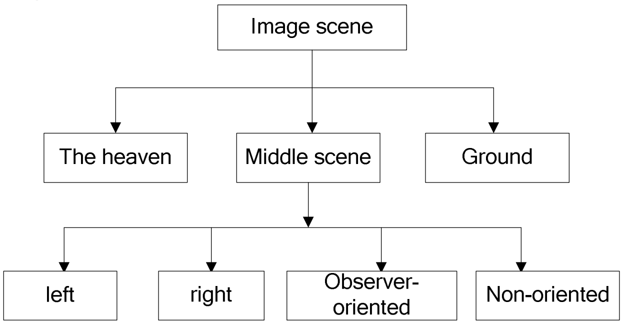

**FIGURE 3: Type of image space layout labeling**

Generally the image should be segmented into different parts to label different areas. Now it is difficult for the existing image segmentation algorithm to realize ideal segmentation, so we propose a multi-segmentation model to minimize segmentation error. First, for an image, we use a diagram-based segmentation algorithm [10] and segment the image by using different parameters to get n different segmentation assumption set $S = (s_1, \cdots, s_n)$ . The optimal segmentation parameters cannot be known in unsupervised image segmentation, so the optimal image segmentation must be assumed and be stored in the set S, or is the combination of different elements in S. even if S has no an ideal segmentation, each element of this set is meaningful for final labeling.

*3.2 Use feature*

Each image block uses the following features in the method of this paper:

Color feature: it includes 3-D HSV mean, 3-D RGB mean, 5-D Hue histogram, 1-D entropy, 3-D saturation histogram and 1-D entropy.

Shape feature: it includes total 1-D pixels/convex closure area;

Texture feature: it includes 12-D DOOG filter output;

Position feature: it includes mean of x and y coordinate of all 2D pixels.

3-D geometric feature: it includes number of 1-D long line, ratio of 1-D approximate parallel long line, distance from1-D disappearing point to image right side, distance from1-D disappearing point to image center, distance from1D disappearing point to image top, 12-D long line directional histogram and 1-D entropy.

At this time, the color, shape and text feature are frequently in different image system. Most images collected by the camera are horizontal, so generally the ground is under the image and the heaven is above the image. The position feature can better distinguish them in our task. Finally 3-D geometric feature is used to determine the orientation of the middle scene. The key issue is how to extract the long line and compute the affine disappearing point.

First, we use canny operator to detect the image edge and extract a long line. We eliminate the line which length is less than the threshold. We select 5% image width in the test. Finally we compute the direction gradient of each pixel for each independent line and represent this line with a 16-D histogram, but if over 90% pixels of a line will be within a dimension in the histogram, we regard it as the long line of this dimension.

Next we will compute the affine disappearing point by using a classical 3-D perspective camera model. In this model, 3-D space $X = (x, y, z, 1)^T$ is mapped to 2-D graphic space $X = (x, y, 1)^T$ . This mapping meets the relation $\lambda x = pgX$ . $p = [l_{3\times3}, 0] \in R^{3\times4}$ is the mapping matrix, $g = (R, T) \in SE(3)$ is rigid transformation represented as a $4\times4$ matrix. $\lambda$ is the scale factor of unknown point *X* on depth *Z*. The line identified with the point $x_1$ and $x_2$ in the image can be represented with a normal plane which is orthogonal to the line l. l is crossing line between the plane passing the mapping center and image plane, namely $l = x_1 \times x_2 = \hat{x}_1 x_2$ , $\hat{x}_1$ is the tilted symmetric matrix related to $x = [x_1, x_2, x_3]^T$ . The public normal place of two lines in the image is $v = l_1 \times l_2 = \hat{l}_1 l_2$ . Given a group of lines, our public disappearing point can be obtained by solving a linear least squares estimation.

$$\min_{v} \sum_{i=1}^{n} (l_i^T v)^2$$

This equation is consistent with solution of $min_v \| Av \|^2$ . Different columns of the matrix $A \in R^{n\times3}$ are the line $l_i$ with same disappearing point direction. In practices, we use lines with similar directional histogram to compose *A* matrix. In addition, the reference [11] describes the method to solve this equation. The figure 4 shows estimation of the long line detection and affine disappearing point. The lines with same color have similar directional histogram.

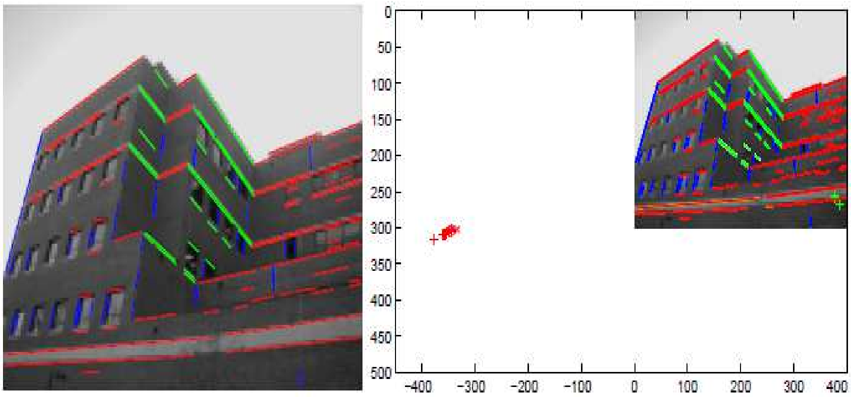

**FIGURE4. Long line and affine disappearing point**

*3.3 Labeling algorithm*

The space layout label is regarded as multi-label problem. Assuming $C=\{c_1,\cdots,c_7\}$ is the label set, $S^t \in S$ is the segmentation assumption and $H^t=\{h_1^t,\cdots,h_m^t\}$ is the image block set in $S^t$. For one image block $h_a^t$ in the segmentation assumption $S^t$, its label vector $l_a^t$ is a 7D vector. $I^{th}$ dimension indicates probability of this image block with label $C_i$ and is represented with [0, 1]. One pixel p of the image belongs to different images in each segmentation assumption. We indicate the image block set covering the pixel *p* in *S* by using $H_p=\{h_p^1,\cdots,h_p^n\}$. $L_p=\{l_p^1,\cdots,l_p^n\}$ is the corresponding label vector set, we can represent label vector of the pixel *p* with $L_p$, namely:

$$L_p=\sum_{i=1}^{n}\pi_i l_p^i$$

$\pi_i$ is the weight factor and is used to represent weight relation between the pixel label vector and the label vectors of different image blocks with this pixel.

It shows that the core of the algorithm is to transform to multi-labeling problem of each image block in the image segmentation assumption set. In practical research, we download 200 training images from Internet, including indoor and outdoor scenarios, segment images into 10 different segmentation assumptions, manually label each image block of each segmentation assumption, and perform iterative training on the labeling set by using *Adaboost*. In the test, we classify the image blocks with unknown labels by using this model and finally process the label vector of each pixel in the images by weighting. In addition, we select 10 images, manually label each pixel, select 100 pixels as the sample T, and get $\pi_i$ by solving a least squares estimation.

$$min\sum_{p\in T}\sum_{i\in[1,10]}(L_p-\pi_i l_p^i)^2,\ s,t,\pi_i\ge 0,\sum_{i}\pi_i=1$$

The figure 5 displays the confidence diagram of different labels by space layout labeling of a test image. The results with ground label indicate the road segmentation.

*3.4 Fusion of space layout and object depth*

After the space layout label of the image is obtained, it indicates that the road segmentation task is roughly completed. To establish more accurate image 3-D spatial model, we combine the space layout label with the depth by using a SVM. In training, we represent each pixel of images with a 8-D vector. 1-7 dimensions indicate the confidence of labels in the space layout labeling. 8th dimension is the estimated depth. In the test, we manually label the obstacle area, use the pixels in the obstacle area as the positive samples, and use pixels in other areas as the negative sample to train a linear core SVM. In the test, we classify each pixel in the images by using the trained SVM and finally determine if each pixel is within the obstacle area [12$^]$.

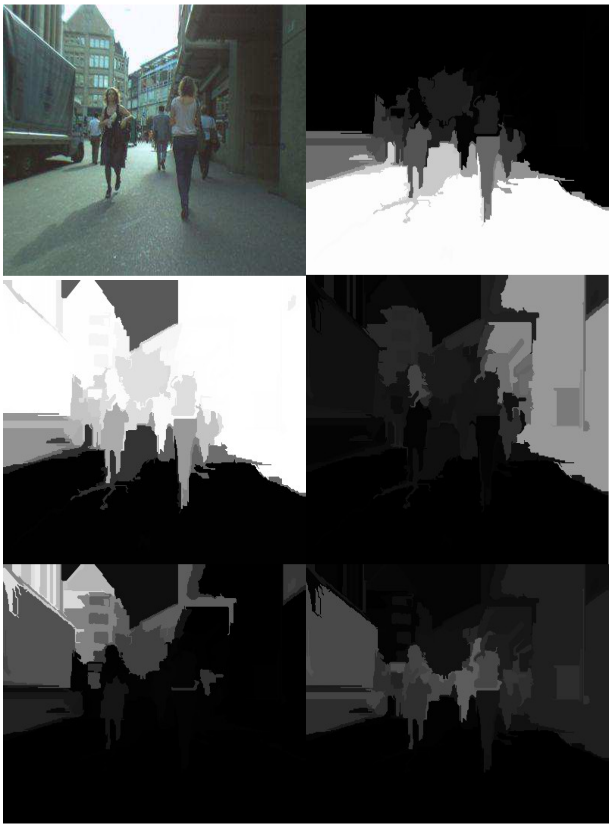

**FIGURE 5：Label confidence diagram of partial space labels**
**(Left up: original diagram. Right up: ground. Left middle: middle scene. Right middle: left object. Left down: right object. Right down: observer-oriented object)**

## EXPERIMENTAL SECTION

To evaluate our method, we test this method by using the ETH obstacle detection data set [13]. This data set includes several continuous video from different street scenarios. Generally the obstacles include pedestrians and vehicles. We select 100 images from the video and manually label obstacles area in 90 images. We train them by using SVM. Obstacles remote from the observer have few influences on the algorithm in partial images, so we only label the main obstacles in the images.in training of SVM and randomly select 1000 positive sample pixels and 1000 negative sample pixels. 50 images are used for test. In evaluation test, we compute the accuracy rate by using $R = \frac{p \cap G}{p \cup G}$. *P* is the image pixels predicted as the obstacles. *G* is the image pixels manually labeled as the obstacle. The table 1 is the final evaluation test results. It shows that the image 3-D information can be mined to most extent by combining the depth with the space layout.

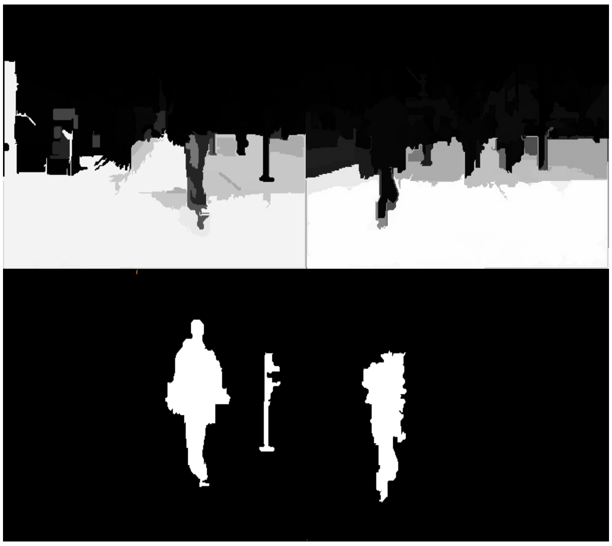

**FIGURE 6 Obstacle detection and road segmentation results obtained by combination of space layout labeling and depth (Up: original diagram. Middle: road segmentation. Down: obstacle detection)**

The experimental results indicate that our scheme has higher accuracy rate for road detection and can recognize the main obstacles in the image in the obstacle detection task. In addition, we further evaluate and test the scheme by using depth information, space layout information and their combination. The figure 6 lists partial test results, which include the obstacle detection and road segmentation results

**TABLE 1: Accuracy rate of obstacle detection**

| **Method** | **Accuracy rate** |
|---|---|
| Depth+SVM | 62.7% |
| space layout labeling +SVM | 68.2% |
| Depth+ space layout labeling +SVM | 79.3% |

## CONCLUSION

In this paper, we establish 3-D spatial model via space layout labeling and depth estimation of the images. The experiment proves that this model can better describe the image scenario, so monocular 3-D reconstruction is qualified for obstacle recognition and road detection. The weakness of this algorithm is difficulty in detection of a long line under fully natural environment, which will affect classification based on SVM later and reduce accuracy rate. We will further improve feature extraction method under the complicated environment, further optimize monocular 3-D spatial model, and enhance its robustness and applicable cases in future.

## REFERENCES

[1] Liu Yuqin, Liu Jingwen. Application of ultrasonic range finder for mobile robotic obstacle avoidance[J]. *Chinese Journal of Scientific Instrument,* **2006**,27:1559- 1560

[2] Li Yunchong, He Kezhong. A Novel Obstacle Avoidance and Navigation Method for Outdoor Mobile Robot Based on Laser Radar [J]. *Robot*, **2006**,28(3):275-278

[3] O hya A, Kosaka A.Vision-based Navigation by a Mobile Robot with Obstacle using Single-camera Vision and Ultrasonic Sensing[J]. *IEEE Trans. On Robotics and Automation*, **1998**, 14(6):969-978

[4] Tan Lei, Wang Yaonan, Shen Chunsheng and Zhou Yuanli. Research on Monocular Vision-based Obstacle Detection for Indoor Mobile Robots [J]. *Computer Measurement and control*, **2010**,18(12):2727-2729

[5] He shaojia, Liu Ziyang, Shi Jianqing. Obstacle detection of indoor robots based on monocular vision[J〕.*Journal of Computer Applications*, **2012**.32(9):2556- 2559.

[6] Du Xin, Zhou Wei, Zhu Yunfang. Obstacle detection based on single view vision[J]. *Journal of Zhejiang University.* **2008**.42(6):913-917.

[7] Liu Wei, Yu Hongfei, Yang Heng. A New Method for Generalized Obstacle Detection Based on Monocular[J]. *Acta Electronica Sinica*.**2011**.39(8):1793-1799.

[8] A. Saxena, S. Chung, A. Ng. Learning Depth from Single Monocular Images[J].*In Proc. Neuron Information Processing Systems*.**2005**.

[9] E. Davies. Machine Vision: Theory, Algorithms, Practicalities 2nd edition[M]. San Diego: Academic Press,

**2005**.

[10] P. Felzenszwalb, D. Huttenlocher. Efficient graph based image segmentation [J]. *International Journal of Computer Vision*. **2004**. 59(2),

[11] K.anatani. Geometric Computation for Machine Vision[M]. Oxford: Oxford Science Publications, **1993**.

[12] Wang Zhengyou Huang Longhua. Image quality assement method based on contrast sensitivity[J]. *Computer applications,* **2006**, Vol.26, 1857-1859

[13] A. Ess, B. Leibe, K. Schindle, L. van Gool. A Mobile Vision System for Robust Multi-Person Tracking [C]. IEEE Conference on Computer Vision and Pattern Recognition, **2008**.