An elevator target tracking method based on kinect

Jian Liu1, Xiaoyu Wang1, Yuanwei Qi1, Ling Chang2, & Rui Sun3

1Shenyang Jianzhu University,Faculty of Information and Control Engineering,Shenyang,Liaoning,110168, China

2Shenyang Urban Construction University,Faculty of Information and Control Engineering,Shenyang,Liaoning,110167,China
3LingyuanUrbanConstructionarchives,Lingyuan,Liaoning,122500,China

Keywords: Target tracking;Kinect; Elevator; Depth information.

Abstract. In order to calculate the number of passengers in the elevator, an elevator target tracking method based on kinect is proposed.When people use the elevator, the person face the camera from into the elevator to leave the elevator back to the camera, which will create some difficulties for detecting face. So the depth of information provided by the kinect somatosensory camera to track when person enter to elevator in this paper. The experimental results show that the depth of information will not be affected by external light conditions, environmental changes and shadows and other unfavorable factors, and be better to track target detection.

1 Introduction

With the need for technological development and practical application, the depth information is increasingly being valued by scholars and studied,especially in areas such does not allow for the measurement of the contact is to get a wide range of applications. Therefore, measuring the depth of information technology and equipment to be shipped out.Based on the depth of the structure of the light collection device Kinect imaging technology,it is possible to provide depth information of the target, the target is the distance of the camera, it is possible to target better tracking.

2 Kinect principle

Irradiating laser light in a space, and the entire space will be marked. Within this space into an object, as long as the speckle image above to see the object can know the location of the object. Roughly calibration method is mainly taken at some distance on a reference surface and the reference plane speckle images recorded. In the measurement,the test space captured speckle image, the speckle image and the recording of this speckle image can be obtained sequentially made a series of mutual correlation operation of the image, the peak in the correlation of the image in the object space exists. After calculating these peaks together, and you will get a three-dimensional shape of the entire scene.

Signals can often be expressed as a linear combination of some orthogonal signal when they are processed. For sparse representation theory, conventional quadrature base is replaced by dictionary which should contain information structure of expressed signal as much as possible. We use the selected dictionary to reconstruct the signal, Reconstruction is essentially the process of approximation of signals with noise, reconstructed signals also remove noise.

© 2015. The authors - Published by Atlantis Press

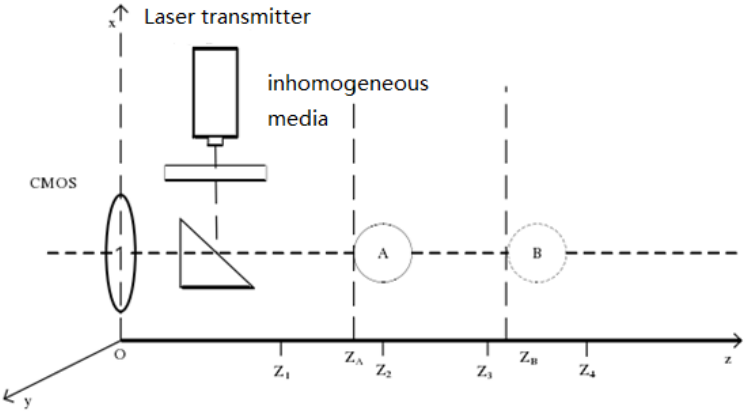

Figure.1: Construction of imaging system

Depth image coding technique is the use of light generated by following these steps:

Step 1: A laser emitter emitting a laser speckle generated diffraction image on the position Z1, Z2, Z3, Z4, and then record these speckle images.

Step 2: A stationary placed in a space object or objects in constant motion, as shown in object A and object B. After laser irradiation, the object A, B surface to form a new speckle, speckle image captured by the CMOS sensor chip.

Step 3: All images speckle image and Step 2 acquired a record of each other in turn calculated the correlation coefficient. To choose the largest correlation coefficient of the reference image, which means that the position of the object is most likely located.

Step 4: According to the calibration relationship between the position of the laser transmitter determines the reference image, can be obtained by calculating the geometric relationship between the object distance to the laser transmitter.

3 Structure learning dictionary

Learning process[4]of Dictionary generally includes two parts: sparse coding and dictionary-updating. First, we need training sample and an original dictionary, then obtain sparse matrix by sparse solving method; Then, update the dictionary according to the learning rule, repeat the above process until we get a satisfactory dictionary.

3.1 Sparse coding

The essence of sparse coding problem is the linear least squares problem of the L1 norm regularization. LARS-Lasso[5] algorithm can solve this problem very well, This algorithm also has a high level of accuracy and robustness.

3.2 Dictionary updating

Initialization: Constructing initial dictionary D_0 and providing training set $w = \{w_1, w_2, ...w_{n-1}\}$, while setting the learning rules $A' = f(A, \alpha)$ and the loop closing conditions T.

Training update :

（1）Order $k = 0$,And assuming that $D = D_0$,

（2）When the loop closing conditions are not met, Execute (3); otherwise do(6).

（3）Sparse coding: Solving equation (3) to obtain $\alpha_k = \arg\min_{\alpha} \| \alpha \|_0$ s.t. $\| w_k - D_k \alpha \| \leq \varepsilon$.

（4）Dictionary updating: $D_{k+1} = f(D_k, \alpha_k)$.

（5）$k = k + 1$,then execute (2) and make judgment.

（6）Learning dictionary $D = D_k$.

Through the above cycle, it is possible to obtain learning dictionary contains a variety of signal component. Using the dictionary and the sparse representation $\hat{\alpha}$ for the signal w which to be processed, we are able to reconstruct the Approximate signal of w.

4 Kinect depth information filtering

When the signal processing, the signal can often be expressed as a linear combination of some basic signal or function. For example, a single signal can represent a combination of a series of sinusoidal signals or become cosine signal. These sine and cosine signals are usually orthogonal.

$$b = \sum_{i=1}^{n} x_i \alpha_i = Ax \tag{1}$$

Among them, it is the coefficient of linear expression vector matrix, as a group-based n-dimensional space. For sparse representation theory, we use the dictionary instead of the traditional orthogonal basis, and the dictionary used to include all of the information structure is expressed as a signal. The use of the selected signal to reconstruct the dictionary, in essence, the signal reconstruction process conducted with noise approximation process, signal reconstruction after it removes noise.

Definition: The set of the original signal and the noise signal composed by approaching:

$$w = w_m + w_r = D\alpha + w_r \tag{2}$$

From the perspective of sparse, hoping to make the case for the smallest, the most sparse solution. Thus, the optimization function can be constructed as follows:

$$\hat{\alpha} = \min \ \| \alpha \|_0 \ s.t. \| w - D\alpha \| \le \varepsilon \tag{3}$$

Among which, $\hat{\alpha}$ is sparse representation for w; $\| \alpha \|_0$ is the norm of L_0, says the number of elements is not 0; ε is error tolerance. We can reconstruct out the signal after eliminating the noise signal in the use of sparse representation and the dictionary. To construct a good dictionary, at present there are many scholars have proposed many effective sparse decomposition algorithm [3], the main have MP, OMP and BP.

5 Kinect simulation figure

Using the Kinect for experiment. The device was based on Windows7 operating platform, called for v1.5 SDK version, it could read the depth information of the image, as shown in figure 1, the left for the real images, the right to read the depth information of the image.

5.1 Reading the depth information

Because of the close kinect object detection accuracy is relatively high, the test for continuous fluctuation denoising facial features. As shown in Figure 4. In Figure 4, left for the depth of the image, we select the characters face de-noising; top right picture shows the depth of simulation map; the lower left is a side view of the depth of the simulation diagram, right picture shows a top view of the depth chart simulation. Region identified from the figure it can be seen, there are a lot of depth information of the face area of the peak and the lack of information.

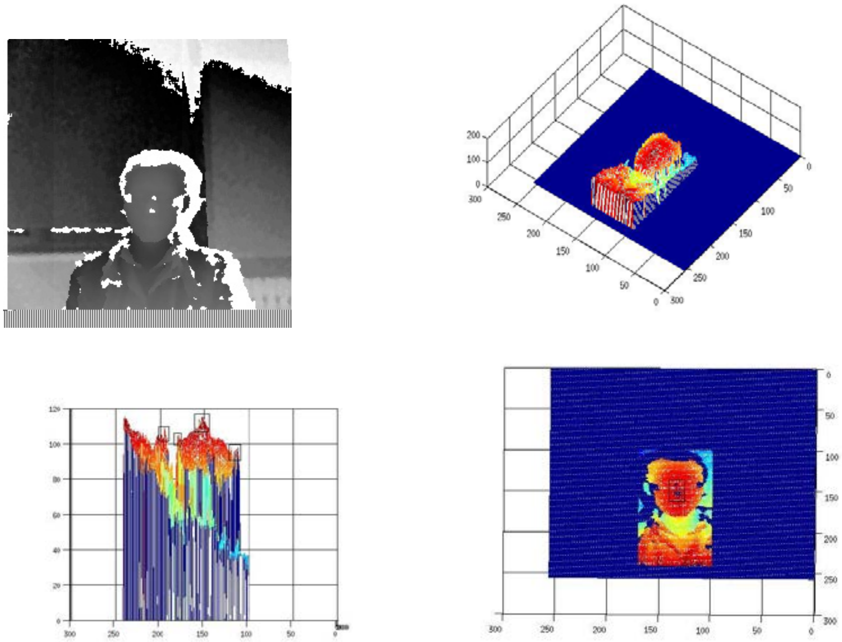

Figure.2: Schematic diagram of approximate image

5.2 Depth information denoising

Constituted by a plurality of two-dimensional normal distribution overlay depth training samples, can find a dictionary to learn the needed training samples below shows the simulation diagram.

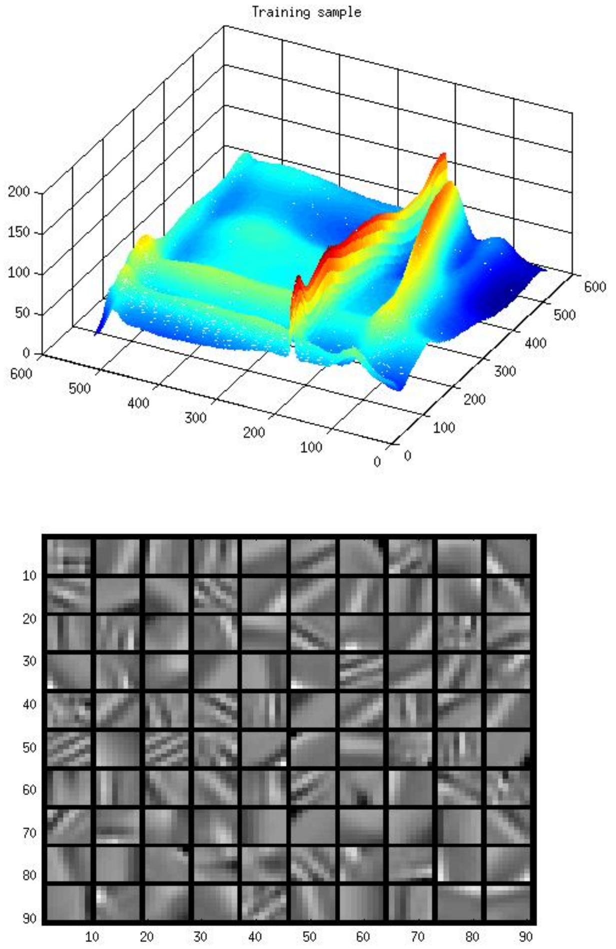

Figure.3: Schematic diagram of training samples and dictionary

Use kinect depth information can be obtained with noise in the form of a 512 × 512 matrix.By studying the dictionary and noisy depth information obtained, and then get noisy sparse representation of information, then the information reconstruction, get close to the noisy signal information, as shown below.

5.3 Comparison of denoising effect

In order to make a good comparison of the effectiveness of the methods mentioned in this paper and several common denoising algorithms in depth information denoising aspect, this paper use gaussian filtering method and mean filtering method for denoising depth information simulation. The following uses peak signal to noise ratio (PSNR), mean square error (MSE)[6], the correlation coefficient (CORR) and other common objective evaluation standard for quantitative comparison.

For depth information reconstruction after denoising, the objective evaluation index calculation results are shown in table 1. Experimental data show that method mentioned in this article has the highest PSNR, the lowest MSE and CORR in the same circumstances for the denoising facial depth information, so the effect of this new method is obvious.

Algorithm	Depth Information512*512		
	PSNR	MSE	CORR
Sparse representation	23.6943	0.0045	0.8019
Gaussian filtering	20.4794	0.0048	0.9032
mean filtering	19.2986	0.0049	0.8947

Table 1: Algorithm comparison of evaluation indexes

6 Elevator target tracking

Because kinect devices were used in the elevator traffic flow monitoring, the use of the video stream sequence, so the need for each frame face detection, after detecting a face to face in the center position for positioning. In the same frame depth image of a face is detected, corresponding to the color images to detect the position of the face of the same, are averaged depth information.

In face of the same depth images, corresponding to detect human faces in color images of the same position, in calculating the average of the depth information. In the next frame video streaming, and expand on 1.5 times the size of a frame to determine the face position, in the area calculation and a frame on the face in different regions in the same depth of average size, and comparing with the desires of average, to draw a frame on the average value of minimum difference of the area, this area is that the frame the location of the face.

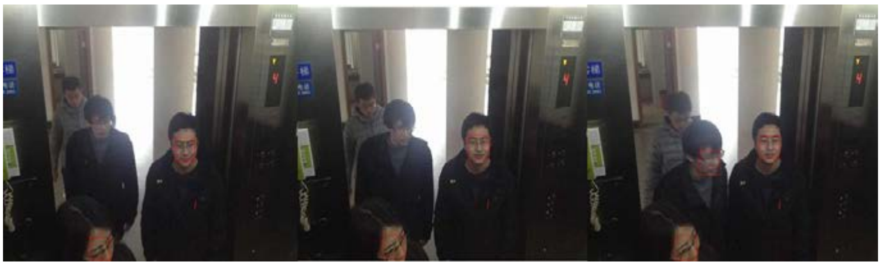

Figure.4: Schematic diagram of face detection in the first layer

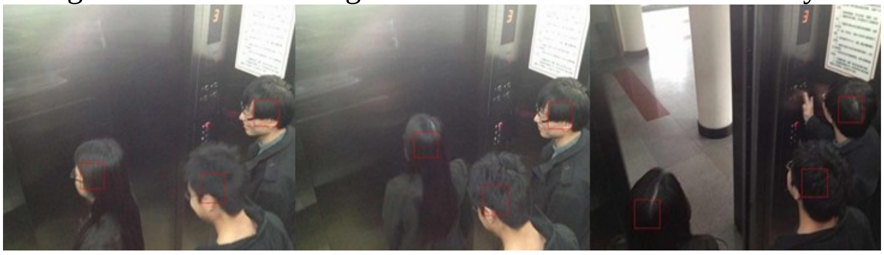

Figure.5: Schematic diagram of head tracking in the elevator car

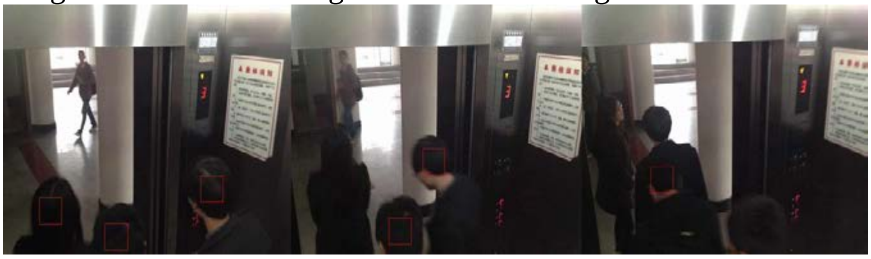

Figure.6: Schematic diagram of overing head tracking

Through the above steps, we can statistics the number of passengers in and out of each floor and every 5 min in and out of the total number of each floor, by artificial statistics compared with algorithm of statistical analysis, statistical process due to the tracking algorithm drift leads to individual tracking data is less than the artificial statistics, but for each group of data statistics algorithm is close to artificial statistical data, the accuracy of the statistics prove algorithm.

7 Summary

This paper proposes the use of statistical methods used to train the elevator and elevator traffic pattern recognition system, then the elevator traffic pattern recognition. Tests proved the depth of information use statistical methods to accurately elevator traffic flow data, and can identify a good elevator traffic pattern, which shows the effectiveness and feasibility of the method used. In short, the use of computer vision for target localization and tracking statistics elevator traffic flow statistics in the field of a new attempt, a new statistical method for solving ideas elevator traffic flow data.

References

[1] Gao.Enyang, Liu. Weijun,Wang. Tianran.Interactive mesh segmentation based on graph Laplacian[C].Applied Mechanics and Materials. Switzerland：Technology for Manufacturing Systems，2011：1535-1540.

[2] Ming-Hsuan Yang, David J Kriegman. Detecting Faces in Images: ASurvey[J]. IEEE Trans. Pattern Analysis and Machine Intelligence,2002,24(1):34-58.

[3] Kwang In Kim,et al.Face Recognition Using Support Vector Machines With Local Correlation Kernels[J].International Journal of Pattern Recognition and Artificial Intelligence, 2002,16(1):97-111.

[4] Lee S. Depth camera image processing and applications[C].2012 19th IEEE International Conference on linage Processing. IEEE, 2012: 545-548.

[5] B.Leibe,A.Leonardis,and B.Schiele.Robust object detection with interleaved categorization and segmentation. IJCV,77(1-3):259–289,2008.

[6] J Miao,et al.A Hierarchical Multiscale and Multiangle System for Human Face Detection in a Complex Background Using Gravity-Center Template[J]. Pattern Recognition,1999,32(7): 1237- 1248.

[7] Tang Yaqin.A Review of Several Image Smoothing Denoising Methods[J].Journal of Southwest University (Natural Science Edition),2009,31(11):125-28.

[8] Wan Xiaohong.An analysis and comparison of the common method for image deniosing[J].Journal of Yuncheng University,2011,29(5):60-62.

[9] Xiao Quan,Ding Xinghao,Wang Shoujue.Image denoising based on adaptive over-complete sparse representation [J].Chinese Journal of Scientific Instrument,2009,30(9):1886-1890.

[10] J.Mairal,F.Bach,J.Ponce,et al.Online Dictionary Learning For Sparse Coding. Proceedings of the 26th Annual International Conference on Machine Leaming, Montreal, Quebec, Canada, 2009:689-696.

[11] M.R.Osborne,B.Presnell,andB.A.Turlach.A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis,20(3):389–403,2000.

[12] Chen Li.Fast Anisotropic Inverse Diffusion For Image Smoothing Denoising[J].Journal of Shantou University(Natural Science Edition),2008, 23(1) : 30 -35.