Available online at www.sciencedirect.com

SciVerse ScienceDirect

AASRI Procedia 3 (2012) 184 – 190

AASRI Procedia

www.elsevier.com/locate/procedia

2012 AASRI Conference on Modelling, Identification and Control

Human Behavior understanding via Top-View vision

Qing Lin[1], CanCan Zhou[*1,2], Shitong Wang[3], Xiaogang Xu[4]

[1,2,4]School of Computer Science and Telecommunications Engineering, Jiangsu University, Zhenjiang, 212013, China
[1]School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, 210094, China
[3]School of Information Technology, Southern Yangtze University, Wuxi 214122, China

Abstract

Aiming to target occlusion problem in complex scenes, human action recognition via a top-view vision is proposed. However, in this view, the human behavior rotated will be mistakenly identified as another behavior. To address this situation, taking into account the rotation invariance of the moments, human static posture are represented by Hu moments, and using the SVM as trainer and classifier. According to the change of coordinate of the binary image centroid, semantic web of dynamic behavior is established. The experimental results show that this method can accurately identify the human dynamic information and has a high recognition rate.

keywords: top-view vision, rotation invariance, Hu moments, SVM, centroid

1. Introduction

Human behavior understanding is very active in the field of computer vision, and with the rapid development of computers and other technology, it has made great progress [1]. In recent years, people have started from theoretical research to practical application scenarios. It is great significance in the real world. Now at many public occasions, the camera is located in the corners or top, so in this paper, based on a positive view would be impractical. At present, most of the behavior studies are based on the corner of the camera, while the top of the camera mostly concentrated on the statistics of the number of people [2,3], and action understanding literature is rare [4].

Typically, vision-based human behavior analysis in general compliances with a few basic processes such as feature extraction, motion characterization, action recognition, high-level behavior understanding and scene understanding [5]. Human action characteristics extracted from image sequences are essential to behavior recognition and behavior understanding. The results based on the positive view and corner view in the complex environment of the room or station will be greatly curtailed by kind of occlusion.

Therefore, a method based on the top-view is proposed. The method has the inherent advantage to solve the occlusion problem. However, in top-view, rotation of people makes the common feature representation is invalid. The translation invariance and rotation invariance of Hu moments are used to represent the gesture feature vector [6]. At the same time more high-level semantic information is drawn according to the change in the coordinates of the centroid.

Corresponding author. Tel.: +86-18796001319.
*E-mail address: youhunzh@163.com，jcs533@qq.com

2212-6716 © 2012 The Authors. Published by Elsevier B.V.
Selection and/or peer review under responsibility of American Applied Science Research Institute
doi:10.1016/j.aasri.2012.11.031

2. Feature Extraction and Movement Feature Representation

2.1 Feature Extraction

The underlying feature extraction is the first step and is also a crucial step in analysis of human action. Whether to extract high-quality moving target is the key to the success of the entire system. Gaussian mixture model (GMM) is commonly used in motion segmentation [7]. It has a good effect on the leaves jitter and water ripples.

K (3~5) Gaussian distribution is indicated the state of each pixel in the image. The probability of Pixel X at time t is:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta(X_t, \vec{\mu}_{i,t}, U_{i,t}) \tag{1}$$

Where K represents the number of Gauss, $\vec{\mu}_{i,t}, U_{i,t}, \omega_{i,t}$ denotes the standard deviation, covariance and weights of the i-th Gaussian at time t, and η represents Gaussian density function.

$$\eta(X_t, \mu_{i,t}, U_{i,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |U_{i,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t-\mu_{i,t})^T U_{i,t}^{-1}(X_t-\mu_{i,t})} \tag{2}$$

If $\left|X_t - \vec{\mu}_{i,t-1}\right| \le \lambda\sigma_{i,t-1}$, you need to update the Gaussian model, otherwise update the minimum right of the Gaussian model, and λ is an empirical value, usually take 2.5 to 3.5.

The update process is as follows:

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha M \tag{3}$$

$$\vec{\mu}_{i,t} = (1-\rho)\vec{\mu}_{i,t-1} + \rho X_t \tag{4}$$

$$\sigma_{i,t}^2 = (1-\rho)\sigma_{i,t-1}^2 + \rho(X_t - \mu_{i,t-1})^T (X_t - \mu_{i,t-1}) \tag{5}$$

$$\rho = \alpha\eta(X_t \mid \vec{\mu}_k, \sigma_k) \tag{6}$$

Among them, α is the learning rate of model, usually take 0.002. ρ is the update rate of model.

In the experiments we found that the scene transient probability is relatively small so model update rate should be taken a smaller one. We sort the weight of GMM, and take the 65% of sum of weights as a useful model. Moving target can be obtained quickly by the above method. However, because of noise, light, and many other factors, the image will appear hollow noise points, shadows, etc. so image post-processing is required. In this paper, the color model is used to calculate the shaded area [8], and the model is described as follows.

Assume that the shadow of the target pixel Shd in the YCbCr space vector represent as Shd(Cx,Cy,Cz). The background pixels 'Back' in the YCbCr space vector is expressed as Back(Bx,By,Bz). Among them, x is the Cb components, y is the Y components and z is the Cr components. The model of YCbCr space is shown in Figure 1.

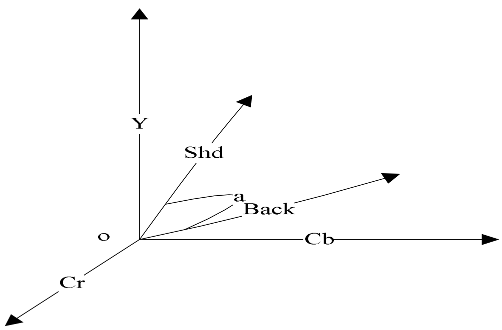

Figure 1 YCbCr space

Figure 2 original picture and processed

Angle a represents the value of the current pixel value relative to the background. The current image and background image pixels of YCbCr known, a is easy to calculate as follows: