# A System for an Anticipative Front Human Following Robot

Weiming Mi, Xiaozhe Wang*
College of Information Science and Engineering,
Northeastern University,
Shenyang, China, 110819
86-18604198395, 86-13614018806
miweiming@foxmail.com;
wangxiaozhe@ise.neu.edu.cn

Ping Ren, Chenyue Hou
College of Information Science and Engineering,
Northeastern University,
Shenyang, China, 110819
86-18660293113, 86-15502446811
rachelrp@163.com;
490069568@qq.com

## ABSTRACT
Human following of mobile robots has been popularized in the scientific field of the service-oriented robots research. However, the behavior of such robots is passive in the most of cases. Compared to a passive following robot, a framework for an anticipative front following robot is proposed which is based on the technology of Kinect camera and the identification of human sensing and motion system. The real-time data is collected by Kinect and applied into the human walking model after median filtering. The human walking model is designed for identifying four basic human walking behaviors accurately. Then a human walking behavior is predicted using the optimized Kalman filtering algorithm proposed specially for the human walking model to predict the next behavior and provide more precise data for the system of robot control, which both could raise the success rate of accurate tracking. The system performance is assessed through experimental comparison of two specific human following behaviors: anticipative front following and passive front following. The results show that the success rate of predicting human walking behavior and following in front achieves 95.2%.

## CCS Concepts
**• Computing methodologies →Vision for robotics**

## Keywords
robots, anticipative front following, Kinect, optimized Kalman filtering algorithm, human walking model.

## 1. INTRODUCTION
The application of service-oriented robots has entered into a prosperous era and has made a significant influence on people's daily life, especially the use of human following robots. Nowadays, most popular type of human following robots which are researched or have been put into market is based on the method of passive following behavior.


*ICAIR and CACRE '16*, July 13 - 15, 2016, Kitakyushu, Japan

DOI: http://dx.doi.org/10.1145/2952744.2952748

For instance, Boston Dynamics has designed a type of robot named BigDog especially for the US military [1]. The BigDog is mainly used to take heavy arms and equipment for US marines and it is a typical passive following robot which need someone to control it. Another example is RobovieII developed in Japan[2], which has been prepared to be put into the market to help the elderly carry things. However, what passive human-following robots could only do is to follow their masters' instructions, which means that such robots could hardly interact with humans actively. This limits the further application of passive human-following robots [3]. Considering the features that an anticipative front following robot should have, we propose a new system for the type of robot with identifying and predicting human walking behavior.

Compared with passive human-following robots, anticipative human-following robots have a higher security of things carried. For example, the intelligent transporting robot, Budgee[4], following after a human, is difficult to protect the goods it takes from being lost or damaged. If a robot follows in front of a human, the human is able to keep an eye on the goods at any time to ensure its intact, which increases the safety of the goods, especially when the goods is precious or its material is fragile. Because the anticipative robot is able to predict human next behavior and it could always be in its master's eyesight, it may receive more popularity than the passive one.

We propose a new control system for front human-following robots. The system is completed in three stages: human walking behavior identification, prediction and front tracking.

## 2. RELATED WORK
The anticipative following robot is mainly controlled by the system software using external devices, such as Kinect camera to control the robot motion. Applying the skeletal data collected from the Kinect into the human walking model, using the judgment procedure of human walking model and the Kalman filtering prediction algorithm, then we will get the robot operation instruction and guide the robot. In this process, the median filtering algorithm is used to improve the accuracy of the initial human bone data. By this way, the impact of noise on the collection of data is reduced, the accuracy of the robot motion behavior is improved.

The median filtering is a nonlinear signal processing technology based on the order statistical theory, which can get rid of image's noise effectively, especially when it is salt-and-pepper noise. In order to apply the technology into 2-D images, standard median filtering is put forward using filter window to remove the impulse noise. Nowadays, a number of researchers have put forward various effective methods based on standard median filtering technology to remove noise in images adapted to images of different characteristics, such as weighted median filtering algorithm[5][6], adaptive median filtering algorithm[7][8], an improved algorithm combined mean filtering and adaptive median filtering[9].

Whether the ability of following the human is good is judged by the robust and accuracy in tracking of a human-following robot. We adopt Kalman filter algorithm to predict the next step that the human would take and to operate tasks to realize the function of anticipative front following. Broida and Chellappa[10] track noisy pixels in images with Kalman filter. There are two main methods to realize point tracking, especially when the size of a target is very small, which means that such target can be simplified as a point. The kernel tracking method based on template and appearance model uses appearance characteristics to build corresponding template in regions of interest and represent the target with histogram. For example, Comaniciu[11] et al. track targets by mean shift algorithm together with color histogram.

As to predicting human walking behavior, there are three main methods: template-based method, probabilistic and statistical method and syntax-based method. Davis et al. try to use power motion image and motion history image to represent a sequence of images and calculate the distance between templates by Mahalanobis Distance[12]. Many methods that adopt K-nearest neighbor (K-NN) classifier actually belong to template-based method. The classification results of this classifier are the most common motion type in its K neighbor training sequences by calculating the distance between image descriptor of observation series and image descriptor of training sequences. For example, Efros et al use method based on optical flow to measure the body in time and space [13], combined with similarity measurement in nearest neighbor method, and to identify human actions in distant. Gorelick with his team detects spatial-temporal features of human body by Poisson distribution[14], with high robustness. Jiang and Lin introduce methods based on learning and matching shape-motion prototype trees[15][16] into recognizing human actions and made k-NN classification. As one of the most commonly used methods for pattern classification, though this method has advantages of low computational complexity, it is difficult to choose time interval.

# 3. CONTROL ALGORITHMS FOR ANTICIPATIVE HUMAN-FOLLOWING ROBOTN

## 3.1 Human walking behavior identification

### *3.1.1 Median filtering algorithm*

We get the skeleton data from depth images by Kinect. Noises in depth images can be disturbed enough to reduce judgment accuracy of human walking posture. Thus it should be the critical step to filter noises when we are getting depth images.

Comparing two depth images, in which one is median filtered and another is without filter, we can figure out that noises can be eliminated a lot after being filtered. And this step can guarantee accuracy when we extract skeleton coordinates from depth images.

(a) Depth images before median filtering

(b) Depth images after median filtering

**Figure 1. Depth images before and after median filtering**

### *3.1.2. Human walking model*

We assume that the skeleton coordinates extracted from depth images which has been filtered is

$$P_H = (X_H, Y_H, Z_H) \tag{1}$$

Where

$$X_H = (x_0, x_1, \cdots, x_{19}) \tag{2}$$

$$Y_H = (y_0, y_1, \cdots, y_{19}) \tag{3}$$

$$Z_H = (z_0, z_1, \cdots, z_{19}) \tag{4}$$

$X_H$、$Y_H$ and $Z_H$ represent the target human's skeletal joints' coordinates.

Computing and analyzing skeleton coordinates and comparing them with the model of human walking postures, we can determine the posture of target human.

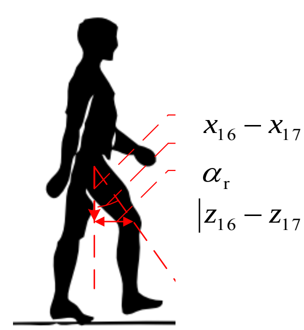



**Figure 2. Human walking model**

Here we only use three pairs of joint points: left hip $(x_{12}, y_{12}, z_{12})$, left knee $(x_{13}, y_{13}, z_{13})$, right hip $(x_{16}, y_{16}, z_{16})$, right knee $(x_{17}, y_{17}, z_{17})$, left shoulder $(x_4, y_4, z_4)$ and right shoulder $(x_8, y_8, z_8)$.

The angle between the vertical and the vector from left hip joint point to left knee is

$$\cos\alpha_l = \frac{|y_{12} - y_{13}|}{\sqrt{(y_{12} - y_{13})^2 + (z_{12} - z_{13})^2}} \tag{5}$$

The angle between the vertical and the vector from right hip joint point to right knee is

$$\cos\alpha_r = \frac{|y_{16} - y_{17}|}{\sqrt{(y_{16} - y_{17})^2 + (z_{16} - z_{17})^2}} \tag{6}$$

Human walking state can be judged by $\alpha_l$ and $\alpha_r$ during human walking.

If the result is that the human is walking, then we judge the way the human walking in; else we can conclude that the human does not move at that moment.

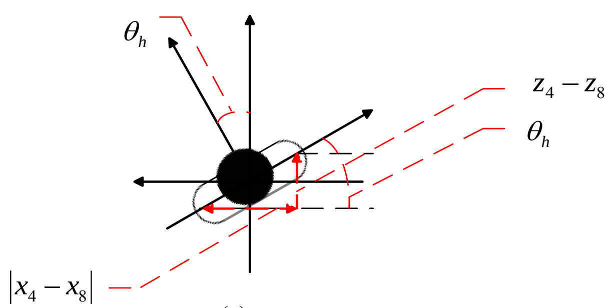
We suppose that the angle between the vectors from the target human's left shoulder to right shoulder and the positive direction of the Kinect is

$$\sin\theta_h = \frac{z_4 - z_8}{\sqrt{(x_4 - x_8)^2 + (z_4 - z_8)^2}} \tag{7}$$

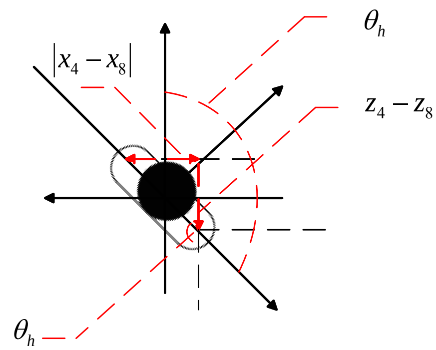
When $|\sin\theta_h| < 0.3$, we suppose that human walks straightly.

When $|\sin\theta_h| > 0.3$, as well as $\theta_h < 90°$, we suppose that human turns left.

When $|\sin\theta_h| > 0.3$, as well as $\theta_h > 90°$, we suppose that human turns right.




(a)




(b)

**Figure 3. Turn left (a) and turn right (b)**

## 3.2 Human Walking Behavior Prediction Based on Optimized Kalman Filtering

There are three common situations during the human-tracking process. When human turns a corner, the robot may be at the human's front side, rear side or the side. If the robot does not correct its direction in time, it will turn out to be bad consequences to keep front-following tracking, even it will lose its target. Shown in Figure 4.

As a result, we adopt Kalman filter algorithm to predict the robot's next position, in order to correct errors during the tracking process.

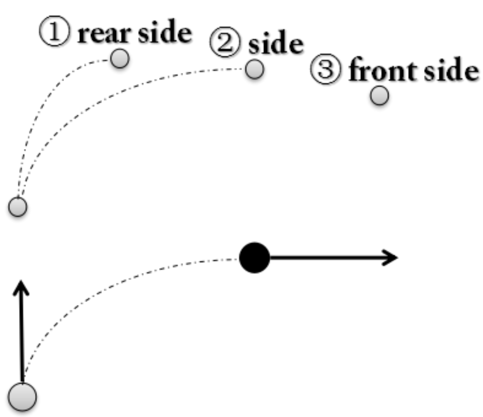



**Figure 4. Situations of losing target**

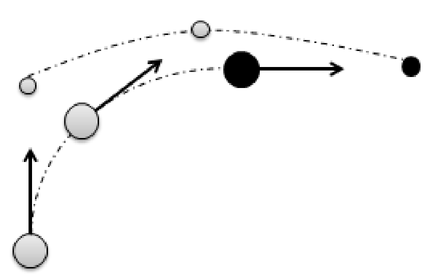


**Figure 5. Correct errors with Kalman filter algorithm**

The coordinate system of Kinect is shown as Figure 6.

We get the coordinates of human's left shoulder $(x_L, y_L, z_L)$ and right shoulder $(x_R, y_R, z_R)$, but here only the coordinates in $xOz$ plane are needed.

Then we can get

$$S_R = (x_R, z_R), S_L = (x_L, z_L)$$

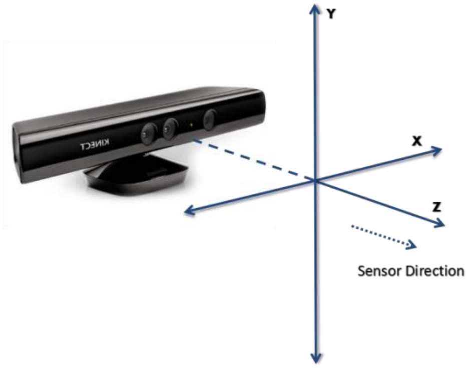



**Figure 6. Establishment of Coordinate System**

The robot is supposed to be right in front of the target human. In other word, it should be at a point which is on the shoulder line bisector.

Thus, the equation which can describe the shoulder line bisector is

$$x = kz + b \tag{8}$$

Where

$$\begin{cases} k = -\dfrac{z_R - z_L}{x_R - x_L} \\ b = \dfrac{(z_R{}^2 - z_L{}^2) + (x_R{}^2 - x_L{}^2)}{2(x_R - x_L)} \end{cases}$$

Expected position of robot is $(x_M, z_M)$ and the vertical distance to the human is $q$. Therefore there is the equation

$$(x_M - x_0)^2 + (z_M - z_0)^2 = q^2 \tag{9}$$

Where

$$\begin{cases} x_0 = \dfrac{x_R + x_L}{2} \\ z_0 = \dfrac{z_R + z_L}{2} \end{cases}$$

$z_M$ is the smaller one of the results in value.

Next, expected position $(x_M, z_M)$ is filtered by Kalman algorithm to predict the next position.

We assume that the true state at time $k$ is evolved from the state at time $(k-1)$ which has been calculated above, then we can substitute values at time $(k-1)$ into the following formula

$$x_k = F_k x_{k-1} + B_k u_k + w_k \tag{10}$$

At time $k$, we have a measurement $z_k$ available, where:

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \tag{11}$$

The equations of predictions are

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{B}_k \mathbf{U}_k \tag{12}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k{}^T + \mathbf{Q}_k \tag{13}$$

And the updated equations are

$$\tilde{\mathbf{y}}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \tag{14}$$

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k{}^T + \mathbf{R}_k \tag{15}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k{}^T \mathbf{S}_k{}^{-1} \tag{16}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \tilde{\mathbf{y}}_k \tag{17}$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \tag{18}$$

# 4. RESULTS AND ANALYSIS

The composition of anticipative following robot is shown as Figure 7. Skeleton data is obtained by Kinect and then matches with human walking model. Kalman filter algorithm is used for prediction here. And then orders for robot are created. Thus, the motion mode of following robot can be modified in time according to the walking mode of the target human. In such pattern, higher accuracy of following behavior can be obtained.

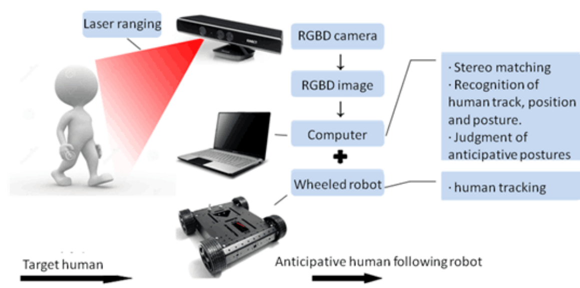



**Figure 7. Construction of the system**

As a 3D camera, Kinect is combined with real-time dynamic capture, image recognition, microphone input, speech recognition, social interaction, and so on. Recently, Kinect has been one of the most commonly used equipment in human-computer interaction research.

Human skeleton is recognized by Kinect in this paper, so that position information can be obtained. In addition, walking posture recognition is based on the human walking model. In terms of motion control, K50 processor is used as core driving device in this paper to control robot tracking behaviors, to make sure that the tracking distance between human and robot appropriate.

In this paper, a complex environment inside of the building is chosen as the experimental scene. The tested walking process of the target person are: holding still, going straight for a distance, doing 90-degree left, going straight for a distance, doing 90-degree right, and finally going straight. The maximum speed of the target person is limited within 2500 mm/s. We use Kinect's depth camera whose image resolution is 640 x 480, and its RGB camera as visual access equipment. The camera frame rate is 30fps. The proposed algorithm is achieved in the Ubuntu system by using C++ language. All experiments in this paper are run on the computer of dual core 1.7GHz Intel CPU.

Figure 8. shows the change of angular velocity of anticipative front following and passive following. The time of rapid change of angular velocity in the anticipative front following mode is the moment to turn after going straight and the moment to go straight after turning. It shows that the robot has a great sensitivity to the turning of the target person. By the figure 8, in the anticipative front following mode, the system can predict the next action, and give the corresponding strategy to meet the demand of the forecast. But in the passive following mode, the angular velocity is gradually changing, because the given strategy can only be based on the distance and the angle difference between the robot's current position and the target person current position. It cannot achieve rapid prediction and change the robot to follow the requirements. Even in the anticipative mode, when the right turn is started, the passive mode is still left turn.

Here in the anticipative following mode, in frame 119, the angular velocity reached the maximum value as 2.520282rad/s and direction is to the right; in the passive following mode, in the frame 114, the angular velocity reached the maximum value as 2.166883rad/s, the direction is to the left.

Figure 9. shows the time-rate change of angular velocity when passive and anticipative front following robots perform. Linear speed of an anticipative front following robot can react quickly to changes that the human makes by predicting the next human walking behavior. However, the linear speed of a passive one changes gradually. There are three periods of time when linear speed of an anticipative front following robot drops significantly due to erroneous judgments of the human's behavior. There are 8 erroneous judgments in the total 170 frames, which means that the accuracy rate is up to 95.3%. In the anticipative front following pattern, except from the large linear speed of the robot at the beginning, the linear speed of the anticipative front following robot reaches its' peak at 2436.723553mm/s in the 127th frame, compared with that of the passive one whose top point is at 2444.824826mm/s in the 132th frame.

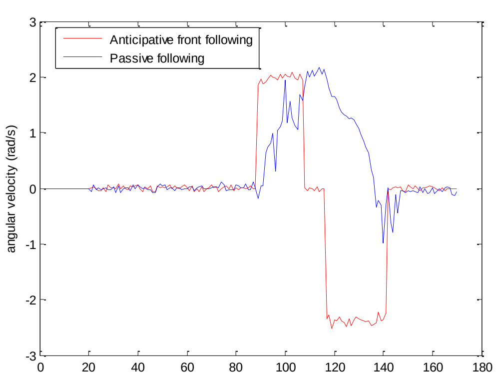



**Figure 8. Comparison of angular velocity**

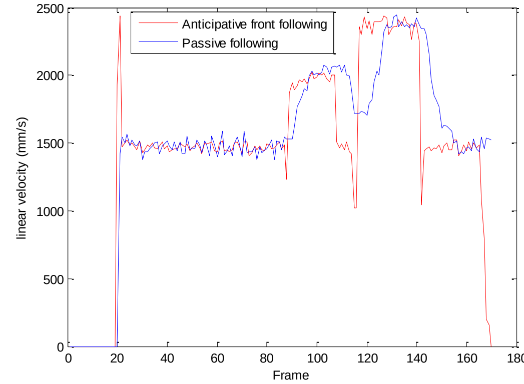
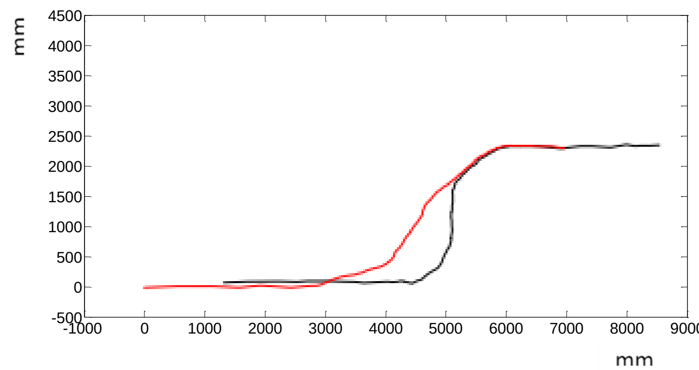



**Figure 9. Comparison of linear speed**

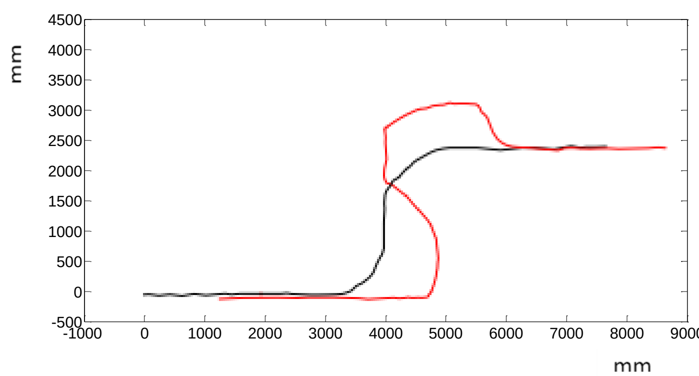
In actual tests, the routes of two different types of robots and the target human are shown in Figure 10.

Red represents the route of robot and black represents the route of the target human.

The horizontal and vertical coordinates represent the length and width of the ground separately and the unit is mm.




(a) Passive front following




(b) Anticipative front following

**Figure 10. Comparison of routes**

As can be seen from Figure 10., passive front following robots track the target human being only by judging the differences of distance and angle between the human and the robot. When people turn right, a passive front following robot is still operating the instruction to turn left due to inertia caused from the virtual spring. As to anticipative front following robot, when the human change the direction, though the robot may run along the original route in a short time, it could adjust the route by turning in a big angle, even the angle could reach 180°. Figure 10. suggests that anticipative front following pattern can change robots route in time by predicting the target human's next behavior. When the similarity between robot's route and human's route is high, it is the time that the robot finishes adjusting its' posture and enters into a stable stage.

# 5. CONCLUSION

At present, because passive robots can only follow instructions rather than cooperate with human beings actively, their application is limited. We propose a system to run anticipative front human-following robots. The process contains identifying human walking behavior through a human walking model, predicting the next walking behavior and tracking. The experimental results show that anticipative front following is able to predict the target's next walking behavior in a very short time and the robot can change its' tracking pattern quickly to secure the integrity of goods carried with a high rate of success. However, due to high sensitivity that the robot react in when people change the walking manner, the route a robot tracks is not as smooth as a passive following robot, which requires a higher standard to carry items.

# 6. REFERENCES


[1] http://news.china.com/jiedu/20140716/

[2] http://jandan.net/2009/12/17/robovieii.html

[3] Hu J S, Wang J J, Ho D M. Design of sensing system and anticipative behavior for human following of mobile robots[J]. Industrial Electronics,IEEE Transactions on, 2014, 61(4): 1916-1927.

[4] http://5elementsrobotics.com/budgee-info/

[5] LI Gang, FAN Rui xia. A New Median Filter Algorithm in Image Tracking Systems[J]. Journal of Beijing Institute of Technology. 2002, 22 (3 ): 376-378.

[6] ZHAO Gao-chang,ZHANG Lei,WU Feng-bo. Application of improved median filtering algorithm to image de-noisi[J]. Journal of Applied Optics, 2011, 32(4): 678-682.

[7] CHANG C C., HSIAG J Y, HSIEH C P "An adaptive median filter for image denoising[J], IEEE Second International Symposium on Intelligent Infarrnation Tcchnolagy Applieatiozi". Washington, DC: IEEE Computer Society, 2008:346-350

[8] WANG Xiao-kai， LI Feng.Improved adaptive median filtering. Computer Engineering and Applications, 2010, 46(3):175-176.

[9] YANG Ming, SONG Li-hua. The application of an improved fast algorithm of median filter on removing image noise[J]. ENGINEERING OF SURVEYING AND MAPPING, 2011, 20(3): 65-69.

[10] Ju S X, Black M J, Yacoob Y. Cardboard people: A parameterized model of articulated image motion[C]//Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on. IEEE, 1996: 38-44.

[11] D. Comaniciu and P. Meer, Mean shift: a robust approach toward feature space analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. 24(5): p. 603-619.

[12] Davis J W, Bobick A E. The representation and recognition of human movement using temporal templates[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 1997: 928-934.

[13] Efros A A, Berg A C, Mori G, et al. Recognizing action at a distance[C]. Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003: 726-733.

[14] Gorelick L, Blank M, Shechtman E, Irani M, Basri R. Actions as space-time shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(12): 2247-2253

[15] Jiang Zhuo-Lin, Lin Zhe, Davis L S. Recognizing Human Actions by Learning and Matching Shape-Motion Prototype Trees. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012, 34:533-547

[16] Jiang Z, Lin Z, Davis L S. A tree-based approach to integrated action localization, recognition and segmentation[M]. Trends and topics in computer vision. Springer Berlin Heidelberg, 2010: 114-127.