TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 12, No. 10, October 2014, pp. 7463 ~ 7470
DOI: 10.11591/telkomnika.v12i8.5581

A Tracking Algorithm of Moving Target in Sports Video

Zhang Hai Tao
Department of Physical Education, Qufu Normal University,
Yantai Road 80, Rizhao City Shandong Province
email: zhanght_2014@163.com

Abstract

Moving target tracking is a core subject in the field of computer vision, its core idea is comprehensive utilization of image processing, video analysis technology, quickly and accurately capture the moving targets. In order to obtain all kinds of technical parameters in the training of athletes, the traditional way is through various sensors used in the athletes. The shortcomings of this method is not put the influence of sensor techniques for athletes in consideration, so it is urgent to have a new technical means to get better movement parameters of the athletes, the traditional target tracking algorithm is difficult to achieve good results, in this study, based on the characteristics of sports video, and it makes some transform to the traditional tracking algorithm, and it proposes a new hybrid tracking algorithm. This algorithm not only can solve the difficult problems of sports video target tracking, and but also can reduce the complexity of the algorithm, through the related analysis and experiments in sports video software, the algorithm can achieve good effect.

Keywords: *tracking algorithm, moving target, sports video, research*

1. Introduction

In the sports video, moving speed of target is faster, and related background is very complex, in generally, it is very difficult to obtain good tracking effect through common target tracking algorithm. So in this paper, it proposes a new global model tracking algorithm, which can solve related problems. Firstly, it adopts the particle filter algorithm to predict the moving targets at the beginning, through similarity comparison of the predicted position of the target with the target model, when the similarity in less than a value, it can be thought that the fundamental changes have taken place in the target movement model, it needs to use a new motion model, when the similarity is greater than a certain value, it can be thought that there is no big changes in the target motion model, it does not need to use new motion model. And then use the mean shift algorithm based on nuclear iteration on the position prediction and thus it can obtain the accurate target location.

At present, this algorithm has been in the practical application of effective inspection, it has received the good effect. The purpose of sports video target tracking is mainly to obtain the track of the moving targets and motion information, at present in this paper, the algorithm has been widely used in our sports training video analysis system.

The global movement is mainly caused by the camera movement in generally. If the camera is in the process of movement, the objects in the images also has their own movements, then the background and foreground in the video sequence has their own movements, the background of the movement is caused by the camera movement, which is known as the global movement. The purpose of global motion estimation is to find out the camera movement law from the video sequence caused by the global movement.

Global motion estimation can be used in the video motion object segmentation such as panorama generation and Psrti coding field, etc. In the video motion object segmentation, global motion estimation is the first step, and then to make estimation of the camera motion between frames and then compensate the frame alignment between the background, and then according to the movement, it can separate the foreground objects and background. The panorama generation is obtained through estimation calculation of global motion between frames and pixels, then according to the movement parameters and through joining together of the adjacent

Received January 5, 2014; Revised July 19, 2014; Accepted August 15, 2014

frames panorama image can be obtained. The Psrti code of MPEG4 is the using of panorama for prediction and compensation, thus it can greatly improve the compression ratio.

Therefore, no matter for the analysis of law of camera motion directly, or analysis of the foreground object motion, the global motion estimation is the basis of these analyses. Global motion parameters estimation method consists of differential method and the method of feature point correspondence, they respectively according to the velocity field or on the image pixel feature points to calculate the relation between the global motion parameters.

In this study, it uses the six parameters model in the modeling of interface scene changes caused by camera motion, and adopts differential method in solving the global motion parameters. When the changes of relative depth of objects in the scene and the camera zoom range is not large, the six parameters of the model can describe the rotation of the camera, lens and displacement movement well. Because in the acquisition of video frame, the adjacent frames can meet the above conditions, so this model can reasonably describe the camera motion between adjacent frames.

2. Camera Model and Camera Calibration

The global motion caused by camera movement on the background can be expressed with 6 parameters motion model just shown as (1):

$$\begin{cases} x_i = ax_i^{'} + by_i^{'} + e \\ y_i = cx_i^{'} + dy_i^{'} + f \end{cases} \quad (1)$$

Where, $p = (x_i + y_i)$ the current coordinates of I_k, $p^{'} = (x_i^{'} + y_i^{'})$ is the adjacent frames, they are corresponding points of $I_{k'}$ and P, $\theta_{k,k'} = (a,b,c,d,e,f)$ is the global parameters, a,b,c,d represent rotation and scaling, respectively, e, f represent the displacement. If in the formula is correct, there should be $I_k(p) = I_{k'}(p^{'})$, so the objective function. is $R(\theta_{k,k'}) = \sum_p \left| I_k(p) - I_{k'}(p^{'}) \right|$. Global motion estimation is to solve undetermined parameters $\theta_{k,k'}$ and get the smallest value of $R(\theta_{k,k'})$. We can use Gauss - Newton or Levengerg- Marquardet nonlinear iteration algorithm. The basic principle of Gauss - Newton iterative method is:

Make the assumptions that for parameters $A_{t,t'}^{k}$, and make the Taylor expansion to objective function R at current point, the type (2) can be obtained.

$$R(A_{t,t'}) \approx R(A^2{}_{t,t'}) + g_k^T(\Delta A_{t,t'}^k) + \frac{1}{2}(\Delta A_{t,t'}^k)^T H_k (\Delta A_{t,t'}^k) \quad (2)$$

Where g_k and H_k represents the gradient and Hessian matrix of $R(A_{t,t'})$ at $A_{t,t'}^k$ respectively.

$$g_k = J_k^T W \gamma_k \quad (3)$$

$$H_k = J_k^T W J_k + \sum_i \gamma_i \omega_i H_{ik} \quad (4)$$

γ_k represent the residuals of $A_{t,t'}^k$, $Jk = \partial\gamma / \partial A_{t,t'}$, W is a diagonal matrix and $W_{ii} = \omega i$, H_{ik} is the Hessian matrix of γ_i , if the value of γ is small, we can think that:

$$H_k \approx Jk^T W_{Jk} \quad (5)$$

$$\partial R(A_{t,t'}) / \partial A_{t,t'} \quad (6)$$

And the in the further we can get:

$$J_k^T W J_k (\Delta A_{t,t'}^k) = -J_k^T W \gamma_k \quad (7)$$

And we can get the resulting incremental $\Delta A_{t,t'}^k$ of $\Delta A_{t,t'}^k$ and then we can find out the next parameter.

$$A_{t,t'}^{k-1} = A_{t,t'}^k - \Delta A_{t,t'}^k \quad (8)$$

Through such iteration, it can gradually reduce the objective function, thus they can obtain the best estimation of the global motion parameters , the throughout iterative process adopts the level of the three-tier pyramid structure in order to improve the efficiency of calculation, the pyramid with [1/4, 1/2, 1/4] filter.

The camera model is the simplification and approximation of optical imaging geometry relationship. Camera model usually consists of a set of parameters, these parameters are called Camera parameters, and the solving process is called camera calibration of camera parameters. Pinhole Model is a kind of ideal Model of the camera. It describes as a central perspective projection imaging process. Perspective projection has the characteristics of small depending almost zero. Parallel straight lines in the space will intersect in the projected image; intersection point is called as the direction of the vanishing point. In addition, in the process of projection point of in the line will keep invariance ratio.

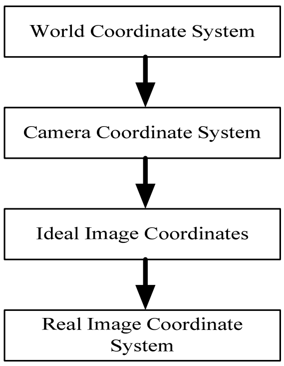

Figure 1. The World Coordinates of the Point Projection to the Image Process

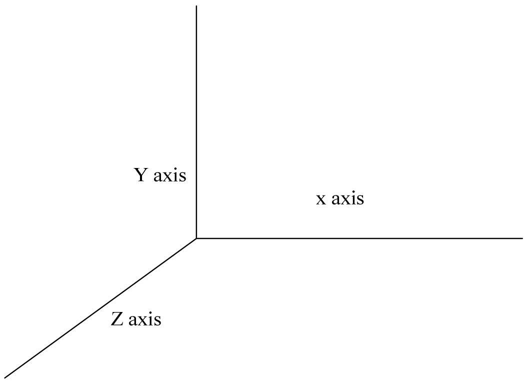

Figure 2. Pinhole Camera Model

Figure 1 presents the simulation projection process of the camera through computer graphics. We call the scene coordinate system as the camera coordinate system, and take the camera as the center, and coordinate system established according to the orientation of the camera is called camera coordinate. And while the image coordinate system is established in two dimensional image coordinate system. The coordinate system and the conventional camera coordinate are alignment. Figure 2 is a pinhole camera model diagram. Figure 4 presents out computer graphics simulation of camera projection process. The scene coordinate system known as the camera coordinate system, established in accordance with the orientation of the camera coordinate system is called camera coordinates and the image coordinate system is established in 2-D image coordinate system, Figure 2 diagram is a pinhole camera model.

If we know in the scene point X_w to the world coordinate system of coordinates $(X_w, Y_w, Z_w)^T$, calculate the point projection also need to know the relative position of the camera coordinate system and world coordinate system (camera position). The relative position can use a rotation matrix R and a translation vector t to express X_w. The projection coordinates can use the formula below:

$$z_c = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f,0,0,0 \\ 0,f,0,0 \\ 0,0,1,0 \end{pmatrix} * \begin{pmatrix} R & t \\ o^t & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (9)$$

So the whole projection process is determined by the projection 3×4 matrix of M, and matrix M can be decomposed into camera external parameters $M_{ex}(R,t)$ which is related to the camera position and $M_m(f_x, f_y, u_0, v_0)$ only related with the camera to the internal structure of camera intrinsic parameters. Camera calibration is in order to determine the parameter matrix, such as the rotation matrix R and related parameters of the displacement vector t.

Above the camera calibration algorithm is not only very complexity, but also needs to solve the camera intrinsic parameters and external parameters, moreover as a result of the real camera especially wide angle lens camera will have a certain amount of distortion compared with the pinhole camera model. So you need to do to make corresponding calibration of the image. Therefore, in this paper, for convenient, it don't use the optimization technique, and also don't introduce nonlinear program, and only adopt linear equation (group) and matrix calculation method. It can make the algorithm more quickly, and the following camera calibration algorithm adopted in this paper is as follows.

First, we select four feature points in the image and on the field model respectively (any of the three feature points not collinear), just as shown in Figure 3:

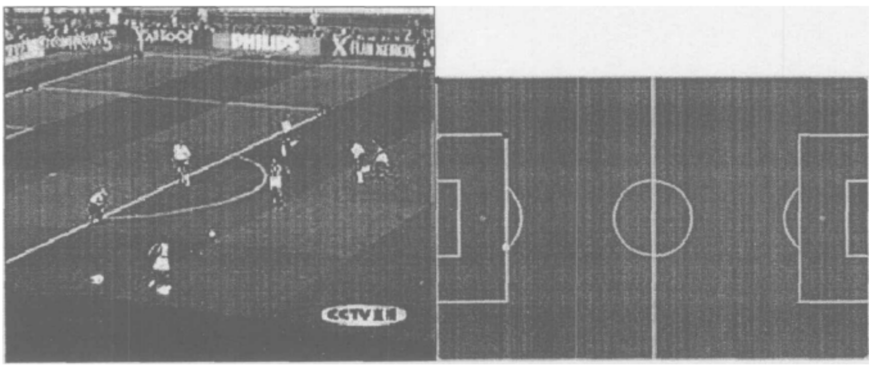

Figure 3. Feature Points on the Model and Image

In the Figure 3, we set the coordinates of the four points on the image are $(x_i, y_i)(i = 0,1,2,3)$, and in the field coordinates of the four points model are:

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} * \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \quad (i = 0,1,2,3) \quad (10)$$

In the process of camera calibration, we can use a set of field and image estimate equation of corresponding points a_{ji}. While in practice, there needs some regularization method to resolve the nonuniqueness problem of type parameter in (10). Usually make element a33 = 1, of course, it also can use other forms of regularization:

From type (10), we can get:

$$\begin{cases} x_i' = \dfrac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}x + a_{33}} \\ y_i' = \dfrac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}x + a_{33}} \end{cases} \tag{11}$$

Through proper algebraic processing, let the a33=l, through each pair of image and corresponding point on the ground model, the of linear equations contains two parameters, expressed in matrix form is as below.

$$\begin{bmatrix} x_i, y_i, 1, 0, 0, 0, -(x_i * x_i'), -(y_i * x_i') \\ 0, 0, 0, x_i, y_i, 1, -(y_i' * x_i), -(y_i' * y_i') \end{bmatrix} * \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \\ a_{31} \\ a_{32} \end{bmatrix} = \begin{bmatrix} x_i' \\ y_i' \end{bmatrix} \tag{12}$$

3. Sports Video Multi-target Trajectory Tracking Algorithm

Multiple target tracking is the key of the current research of computer vision, especially in the human body tracking field, it is a hot research point. The multiple targets tracking algorithm is roughly divided into two categories: the first one is based on the model of multiple targets tracking system; and the second one is based on the multi-source information fusion of multiple targets tracking system. Multiple targets tracking algorithm which is based on the model, and it is mainly use the relationships between the tracked targets, the movement of the multi-objective model is established. And then reuse the corresponding state space search target tracking algorithm, it can achieve the target motion, the algorithm is mainly used in the human body tracking. The multi-source information fusion algorithm of target tracking is often through multiple sensor to realize the information fusion. And then through using neural network and hidden Markov model to realize information fusion. This kind of algorithm is mainly used in radar signal processing, etc.

In the study, the motion estimation algorithm based on particle filter algorithm and video cameras is adopted in the analysis of football, hockey video. The aim is to track athletes running route so as to obtain the movement estimation of the of athletes and sports information such as speed, which can provide the help for coach in tactical analysis.

As the traditional multiple target tracking algorithm, it is often based on the static background, and only can get the speed and trajectory if the target which has relative movement to the camera, it can't get real movement information of target. so the algorithms cannot provide useful information for the coach.

In order to get the motion estimation of the players on the pitch, the first step is to get the trajectory of players in the video, and then reuse the trajectories of camera motion estimation algorithm for camera, through the video image space coordinates with the mapping relationship between real space coordinates, finally get the motion of players on the pitch. The algorithm process algorithm is as follows:

1. Through using of camera calibration algorithm, it can get the mapping relationship between the model and the first video images:

$$\begin{cases} x' = \dfrac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}x + a_{33}} \\ y' = \dfrac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}x + a_{33}} \end{cases} \quad (13)$$

Where the point (x', y') is the coordinates of the field model, the point x (y) is the coordinates of video image in first frame.

2. Through hybrid particle filter and the mean shift tracking algorithm, the coordinates of the player on the current video frame (x_k'', y_k'') can be obtained.

3. Make the assumption that $(x_i, y_i)^T$ is a pixel position in the current frame image, $(x_i, y_i)^T$ is the point of the image location, the relationship between them can be expressed as:

$$\begin{cases} x_i' = (a_0 + a_2 x_i + a_3 y_i) \\ y_i' = (a_1 + a_4 x_i + a_5 y_i) \end{cases} \quad (14)$$

$$\begin{pmatrix} x_i' \\ y_i' \end{pmatrix} = A \begin{pmatrix} x_i \\ y_i \end{pmatrix} + T \quad (15)$$

Among them, $A = \begin{pmatrix} a_2 & a_3 \\ a_4 & a_5 \end{pmatrix}$ represent zoom, rotate and stretching, $T = (a_0, a_1)^T$ represent translational motion. And through applying the global motion estimation algorithm, we can obtain the camera motion parameters $\begin{pmatrix} a_{ko}, a_{k2}, a_{k3} \\ a_{k1}, a_{k4}, a_{k5} \end{pmatrix}$.

4. Make the solution of (x_k'', y_k'') coordinate of tracked target point on the current frame, which is corresponding to the first frame image coordinates $(\overline{x}_k, \overline{y}_k)$.

$$(x_k'', y_k'')^T = \prod_{j=1}^{k} \begin{pmatrix} a_{j0}, a_{j2}, a_{j3} \\ a_{j1}, a_{j4}, a_{j5} \end{pmatrix} \begin{bmatrix} 1 \\ \overline{x}_k \\ \overline{y}_k \end{bmatrix} \quad (16)$$

$$(x_k'', y_k'')^T = A \begin{bmatrix} 1 \\ \overline{x}_k \\ \overline{y}_k \end{bmatrix} \quad (17)$$

Where,

$$A = \prod_{j=1}^{k} \begin{pmatrix} a_{j0}, a_{j2}, a_{j3} \\ a_{j1}, a_{j4}, a_{j5} \end{pmatrix} = \begin{pmatrix} a_{0,} a_2, a_3 \\ a_1, a_4, a_5 \end{pmatrix} \quad (18)$$

$\begin{pmatrix} a_{j0}, a_{j2}, a_{j3} \\ a_{j1}, a_{j4}, a_{j5} \end{pmatrix}$ is the global motion estimation parameters of the video from frame j-1 to frame j .

$$\begin{cases} \overline{x}_k = \dfrac{(a_5 x_k'' + a_1 a_3) - (a_0 a_5 + a_3 y_k'')}{(a_2 a_5 - a_4 a_3)} \\ \overline{y}_k = \dfrac{(a_2 y_k'' + a_0 a_4) - (a_1 a_2 + a_4 x_k'')}{(a_2 a_5 - a_4 a_3)} \end{cases} \quad (19)$$

5. Through the mapping relationship between the ground model and the first frame video image, we can obtain the coordinates of tracked target points (x_k'', y_k''), corresponding coordinates of the point on the model (x_m'', y_m'')

$$\begin{cases} x_m'' = \dfrac{a_{11}\overline{x}_k + a_{12}\overline{y}_k + a_{13}}{a_{31}\overline{x}_k + a_{32}\tilde{y}_k + a_{33}} \\ y_m'' = \dfrac{a_{21}\overline{x}_k + a_{22}\overline{y}_k + a_{23}}{a_{31}\overline{x}_k + a_{32}\overline{y}_k + a_{33}} \end{cases} \quad (20)$$

Below is an experimental result of Brazil vs England 2002 World Cup football game. They are 12932 to 13234 frames of the match video.

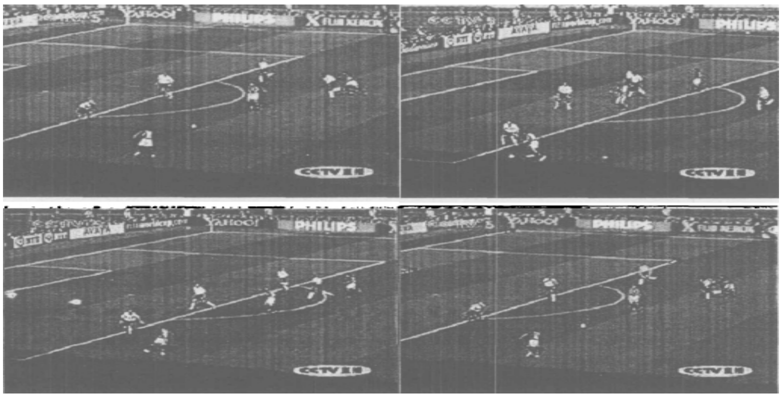

Figure 4. The Image of the Football Match

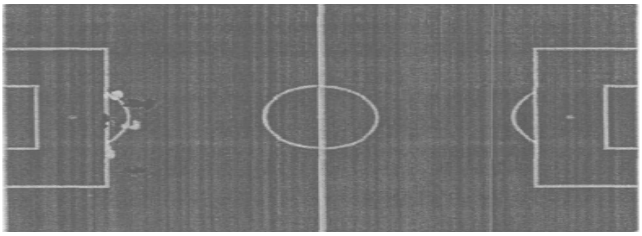

Figure 5. The Trajectory Tracking Results

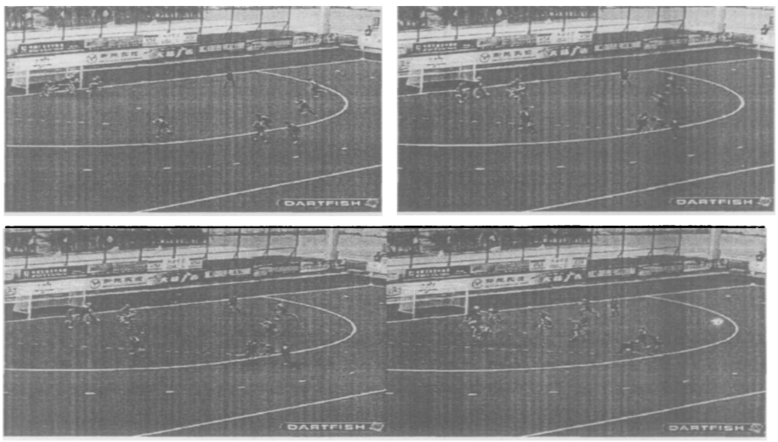

Figure 6. The Image of the Football Match

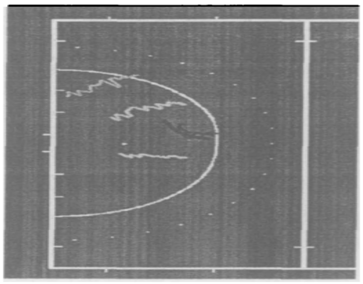

Figure 7. The Tracking Results of Red Team

4. Conclusion

The algorithm combins particle filter and global motion estimation algorithm, and it can solve the problem of multiple target tracking algorithm which only can receive location coordinates relative to the camera coordinate system, and the shortcoming of unable to get the real motion position also can be waked.

The article also is aim at the player position overlapping phenomenon in multiple target tracking, the Bayesian classifier is added, and through using the target movement information, it can preliminary solve the athletes overlapping case, the tracking problem of athletes. From Figures 4-7, it can be see that, the algorithm proposed in this paper, the camera angle is changed, will still be able to accurately track target position on the pitch, so as to solve the multiple target tracking algorithm before only can track moving target relative to the position of the video and can't get the actual location. The camera motion estimation algorithm can be applied to the algorithm in target tracking, thus make the multiple target tracking algorithm can be applied to the dynamic background and solve the dynamic background, target positioning problem at same time.

Acknowledgements

The research is supported by Soft science research project of Shandong province general project (2013 RKB01025)

References

[1] C Wren, Azarbayejani, T Darrell, A Pentland, Pfinder: Real-time Tracking of the HumanBody. *IEEE Trans. PAMI,* 1997; 19(7): 780-785.
[2] T Olson, F Brill. *Moving Object Detection and Event Recognition Algorithms for Smart Cameras*. Proc.DARPA Image Understanding Workshop. 1997.
[3] TN Tan, GD Sullivan, KD Baker. Model based localization and recognition of road vehicles. *Int. J. Comput. Vis*. 1998; 29(1): 22-25.
[4] RT Collins, AJ Lipton, T Kanade. *A System for Video Surveillanceand Monitoring*. Proc.Am.Nuclear Soc. (ANS) English Int. Topical MeetingRobotic and Remote Systems.1999
[5] MJ Swain, DH Ballard. Color indexing, *International Journal of Computer Vision*. 1991; 11-32.
[6] R Kieldsen, JKender. *Finding skin in color images.* 2nd international conference on Automatic Face and Gesture Recognition, Killington,Vermont, USA. 1996.
[7] P Fieguth, D Terzopoulos. *Color-based Tracking of heads and other Mobile objects at video frame rates.* Proc.Of IEEE CVPR. 1997: 21-27.
[8] M Hunke, A Waibel. *Face location and trackingfor human-computer interaction.* Proc of the 28th Asilomar Conf. ON Signals, Sys. and Comp., 1994: 1277-1281.
[9] K Sobottka, I Pitas. *Segmentation and tracking of faces in Color Images.* Proc of the Second Intl.Conf.On AutoFace and Gesture Recognition. 1996: 236-241.
[10] Y Bar-Shalom, T Fortmann. Tracking and Data Association. Academic Press. 1998.
[11] S Julier, J Uhlmann. *A New Extension of the Kalman Filter to Nolinear Systems.* Proc. SPIE. 3068. 1997; 182-193.
[12] Michael Isard, Andrew Blake Condensation-Conditional Density Propagation for Visual Tracking. *International Jounal of Computer Vision.* 1998; 29(1): 5-28.