We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK
CITATION
INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Human Motion Tracking Based on Unscented Kalman Filter in Sports Domain

GuoJun Liu and XiangLong Tang
Harbin Institute of Technology
China

1. Introduction

This work aims to automatically track the movements of ice skaters on a large-scale complex and dynamic rink. From the computer vision point of view, several open challenging problems in the tracking of sports athletes present themselves (Intille & Bobick, 1994): the skaters move rapidly and change direction unpredictably; the estimation of skaters' motions is compounded with non-smooth camera motion; skaters range in size from 30 by 30 pixels to about 70 by 70 pixels, depending on the setting of the camera; competitors flail arms and legs frequently during the match. Additionally, partial occlusions often occur when skaters are close or skaters collide during the overtaking in the short track speed skating games. A one-second-long sequence of a single skater is shown in Fig. 1.
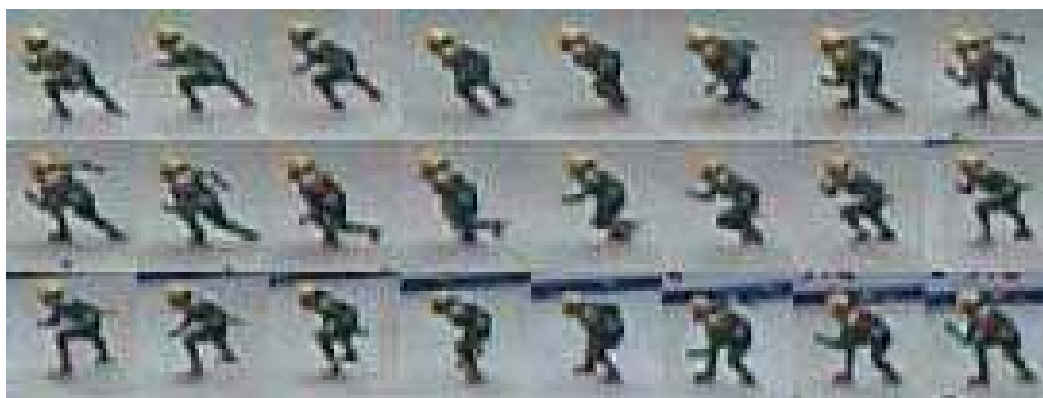

Fig. 1. A one-second-long sequence of a single skater.

There is a substantial literature on tracking for various purposes including video surveillance, smart environments, pedestrian, face tracking, action recognition in sports and many other domains. Haritaoglu et al. (Haritaoglu et al., 2000) use a variation of the background to detect object in the foreground. Rigoll et al. (Rigoll et al., 2004) use stochastic modeling techniques, Pseudo-2D Hidden Markov Models (P2DHMMS) and Kalman filter to estimate the location of a person. In (Ozyildiz et al., 2003; Rasmussen & Hager, 2001; Yilmaz et al., 2004), a technique of fusing multiple cues (e.g. texture, color and shape) is employed to track. The mean shift algorithm (Comaniciu et al., 2003; Yang et al., 2005) has achieved considerable success in object tracking due to its simplicity and robustness, it finds local minima of a similarity measure between the color histograms or kernel density estimates of

the model and the target. Recently, the integration of color distributions into particle filter becomes popular due to the outstanding performance on tracking the non-rigid objects with the non-linear and non-Gaussian motion (Nummiaro et al., 2003). But these well-known traditional tracking algorithms are inadequate for tracking the low-resolution, amorphous, fast erratically moving and colliding skaters in a large-scale, complex and dynamic scene by using a single panning camera.

The remainder of this paper is organized as follows: the next section explores the related work in various other sports fields. Section 3 overview of our system, Section 4 explains how to automatically compute the mappings that transform each frame to the model of the rink and section 5 describes our tracking method which combines the unscented Kalman filter with a hierarchical model based on contextual knowledge. Experiments and results are given in section 6 and the conclusion and discussion in section 7.

2. Related work

As the most popular sport in the world, soccer has enormous potential market value. The related applications, such as broadcast or annotation systems, are receiving more and more attention from computer vision researchers. Misu et al. (Misu et al., 2002) address the problem of tracking soccer players in occlusion by fusing different objects features like textures, color statistics, movement vectors, etc. In (Vandenbroucke et al., 2003), color image segmentation by using pixel classification in an adapted hybrid color space is applied to extract those meaningful regions that represent the players and recognize their team. Needham and Boyel (Needham & Boyel, 2001) describe a CONDENSATION based approach on image sequences obtained from a single fixed camera, Kalman filter and ground plane information are used to improve the prediction of player movement and assist to track occluded players. Junior and Anido (Junior & Anido, 2004) develop a real-time distributed system by using five static cameras to get the whole scene, each player can be tracked individually by an independent module of the system, and each module can apply different tracking techniques depending on specific visual characteristics.

There are many computer vision systems for other sports domains. Pingali et al. (Pingali et al., 1998) develop a real time tracking technology for enhancing the broadcast of tennis matches from stationary cameras. Recently, Yan et al. (Yan et al., 2006) propose a data association algorithm to track a tennis ball in low-quantity tennis video sequences. Pers et al. use two stationary cameras mounted directly above the court and propose a new approach for modeling the radial image distortion more accurately. The tracking algorithm combined with color feature and the template is exploited to track the player, using color feature is to avoid a drift caused by the template tracking. Their systems are applied to many sports domains including squash (Pers et al., 2001), handball (Pers et al., 2002) and basketball (Jug et al., 2003). But there are two limitations in their works: first, the cameras must be placed above the playing court, which is a rigorous condition for regular league or championship matches. Second, how to handle occlusion in tracking process seems not to be solved.

Intille and Bobick (Intille & Bobick, 1994) develop the state-of-art automatic annotation system for American football footage and lay a foundation for research in the automatic annotation of video. In their system, camera motion is to be recovered by using a global model of the football field which consists of some geometrical primitives (e.g. lines) and some features (e.g. number, logo). The "close-world" is defined as "A region of space and

time in which the specific context is adequate to determine all possible objects present in that region", the tracking is performed on those player's pixels extracted from motion blobs by "close-world" analysis. Okuma (Okuma, 2003) develops a hockey annotation system to automatically analyze hockey scenes, track hockey players in these scenes and construct a visual description of these scenes as trajectories of those players. Their system has two components: one is rectification system that transforms the original sequence in broadcast video to the globally consistent map of the hockey rink. The other is a color-based sequential Monte Carlo tracker. Robustness and reliability of this system are shown that the tracker tracks a single target for about 500 frames and the error between the estimated position and the real world position is from 0.3 m to 1 m. Compared with (Okuma, 2003), the framework of our system is similar, but our method is very different from theirs in many aspects such as camera plan, registration and tracking algorithm.
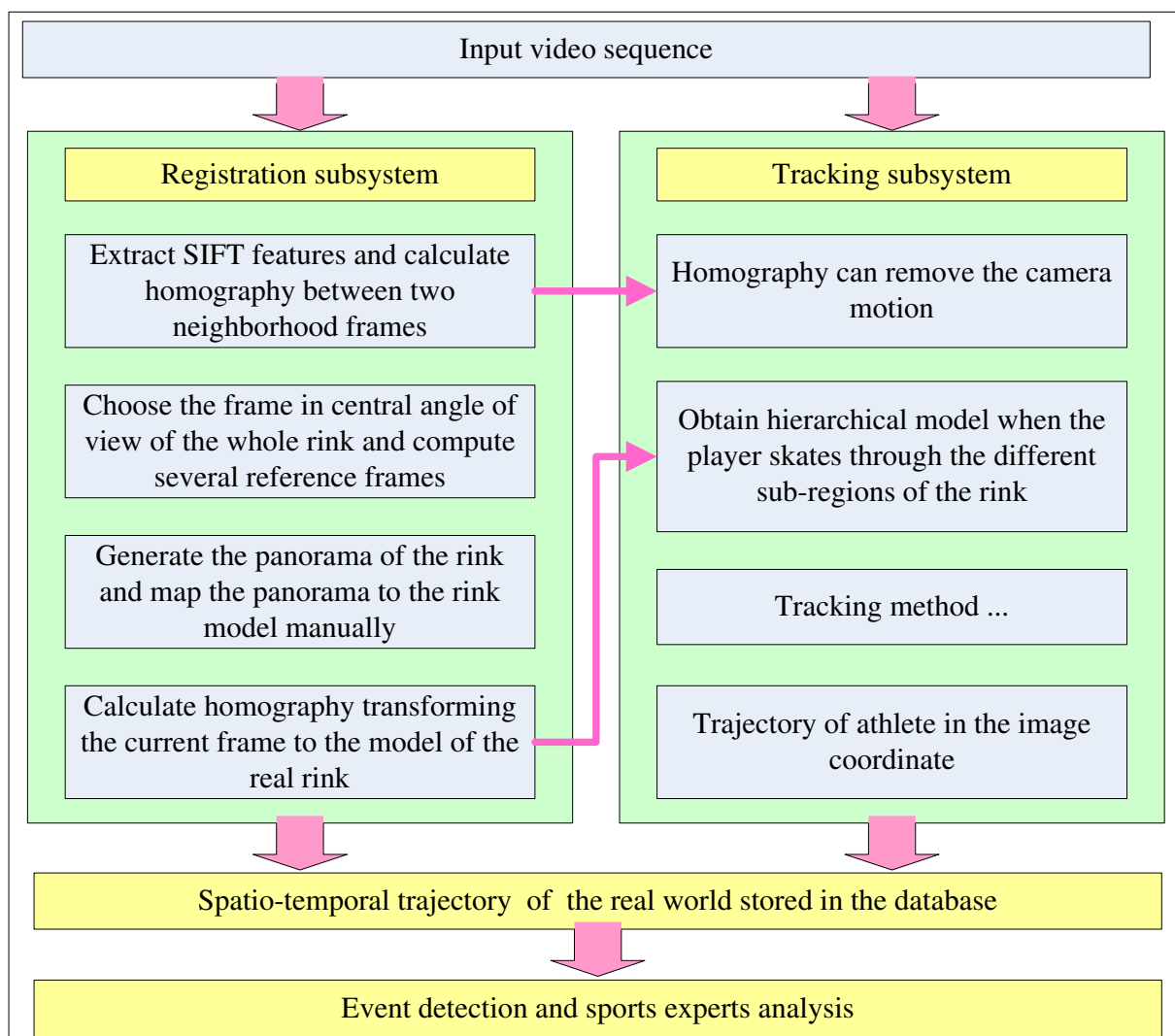
3. Overview of our system

Fig. 2. The architecture of our system.

Our computer vision system aims to automatically track the movements of skaters on a large-scale complex and dynamic rink. Our aim is to exploit it not only in daily training but also in competitions. We used a single panning camera, which was mounted at the top auditorium of the stadium as close as possible to the center in order to reduce the projection error. Due to little texture information on the rink, unlike (Intille & Bobick, 1994; Okuma, 2003), zooming was abandoned, it can make recording the high-speed target more difficult and enlarge the error of lens distortion. Though the camera center moves by a small amount, due to an offset from the camera's optical center, the approximation of pure rotation is indeed sufficient such as proven in (Hayman & Murray, 2003).

The architecture of our system is shown in Fig. 2, it includes two kernel subsystems. Automatic registration affects directly not only the accuracy of the system's output but also the tracking performance, it can provide two kinds of outputs to other modules: (1) The homography between two neighborhood frames can remove the camera motion to improve the precision of the motion prediction in the process of tracking. (2) Another homography transforming the current frame to the model of the real rink is associated with the output of tracking subsystem, such as the skater's spatio-temporal trajectories in the real world, and that can be stored in the database. At the same time, it can be used in the tracking process to obtain a skater's hierarchical model when the skater moves through different sub-regions of the rink. To improve the tracking performance, the proposed tracking subsystem incorporates the hierarchical model based on contextual knowledge and multiple cues into the unscented Kalman filter, the details will be discussed in section 5.

4. Automatic registration

The goal of automatic registration for sports applications is transforming the positions in the video frame to the real-world coordinates or vice versa. Generally, the registration needs many point or line features extracted automatically from images, then matches them and uses the correspondence to calculate the homography. Farin et al. (Farin et al., 2004) use a model of the arrangement of court lines for registration, similar to the one in (Hayet et al., 2004). However, if no obvious lines are in the fields or courts, it can't work well. Okuma et al. (Okuma, 2003) compute the local displacements of image features by using the Kanade-Lucas-Tomasi (KLT) tracker (Shi & Tomasi, 1994; Tomasi & Kanade, 1991) and determine the local matches, but it needs to predict the current camera motion based on the previous camera motion to reduce the amount of local feature motion, if the camera moves rapidly and asymmetrically such as in our application, the worse prediction can lead to the KLT tracker failing.

Therefore, for a rapid camera, the registration faces two big difficult problems: (1) Detect the more distinctive point features and match them better. (2) How to reduce the accumulative registration error for a long image sequence.

4.1 Homography

A homography is a projective transformation which is a nonsingular matrix H with size 3×3. If x and $x^{'}$ are the images of a world point, belonging to a plane, they are related by a matrix H corresponding to that plane: $x' = Hx$, since the matrix H has 8-Degree of Freedom (DOF), 4 point correspondences can determine H. Obviously, with non-perfect data, more points should be used.

In general, the moving targets in the image can decrease the accuracy of the homography since the good features on the moving ones are imperfect. The RANSAC algorithm (Fischler & Bolles, 1981; Hartley & Zisserman, 2000) can pick out the worse features (namely outliers) easily, as shown in Fig. 3. However, it can only assure the computational accuracy between two adjacent frames, not for a long video sequence. How to reduce the accumulative error is the key issue.
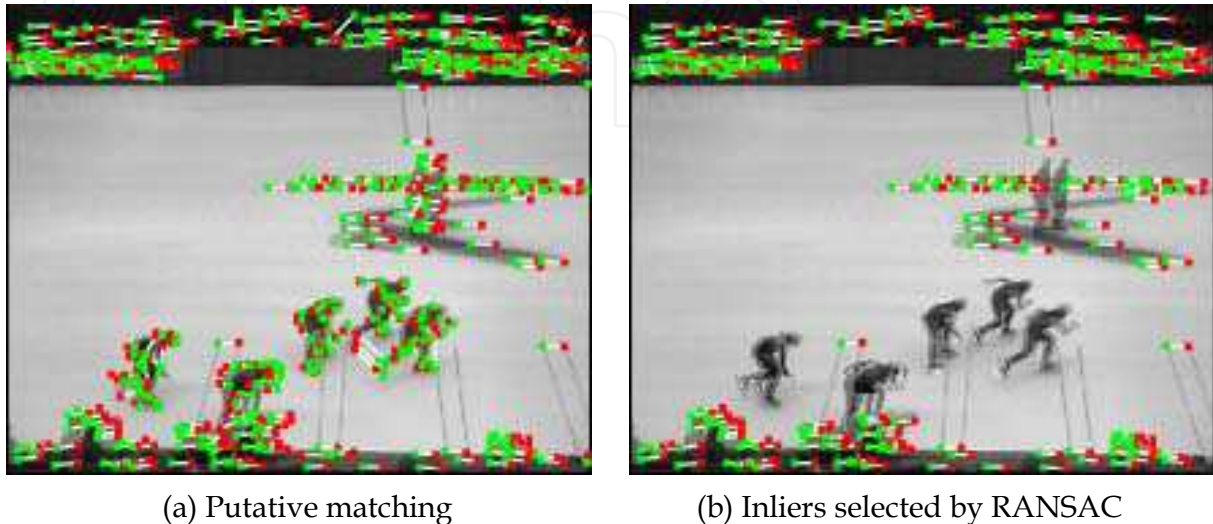

(a) Putative matching   (b) Inliers selected by RANSAC

Fig. 3. The outliers on the moving skaters and the referee are eliminated.

4.2 Selection of point features

Feature extraction and matching determine the computational precision of the homography and the overall reliability of the registration algorithm. Two of the most popular feature extractors are the Harris corner detector (Harris & Stephens, 1988) followed by the Sum of Squared Difference (SSD) matching, and the Kanade-Lucas-Tomasi (KLT) tracker (Shi & Tomasi, 1994; Tomasi & Kanade, 1991). These methods can work well when the baseline is relatively small and the appearance of the features doesn't change too much across the subsequences. A more distinctive feature is desirable for matching two wide baseline images (e.g. a large translation, scaling, or rotation between two frames) such as the current frame and its corresponding reference frame in our application.

Lowe (Lowe, 2004) proposed the Scale Invariant Feature Transformation (SIFT), where a feature's location and scale are determined by extrema of a Difference of Gaussian (DOG) function in the scale space, and its dominant, local image gradient orientation. It is invariant to viewpoint changes, large geometric transformation, changes in illumination, and has been applied in the areas of object recognition (Sivic & Zisserman, 2003) and panorama stitching (Brown & Lowe, 2003). Therefore, it is suited to dealing with the problem of wide baseline matching.

4.3 Registration method for a long image sequence

In order to reduce the accumulation of the errors generated from the set of frame-to-frame homography, many reference frames are introduced and calculated to construct a panoramic image (Brown & Lowe, 2003; Shum & Szeliski, 2000) as illustrated in Fig. 4 and

each frame can be mapped to the most adjacent reference frame. The model of the entire rink is shown at the bottom in Fig. 4, which includes precise measurements of geometrical features: start line, finish line and marking blocks, and all that can be obtained from ISU (International Skating Union, http://www.isu.org/). The corresponding points labeled from 1 to 20 are initialized manually, the detailed procedure is illustrated in Algorithm 1.

Algorithm 1 Registration method for a long image sequence

Input video sequence and perform the following steps:

1. Compute homography $H_{t-1,t}$ between the frame at time $t-1$ and the frame at time t with the RANSAC algorithm (Hartley & Zisserman, 2000).
2. Choose the frame in the central angle of view of the whole rink as a base frame used to construct the panorama.
3. Compute the reference frames which distribute on the both sides of the central frame at regular intervals.
4. Generate the panorama of the rink with all reference frames and calculate homography $H_{ref^i,pano}$ transforming the ith reference frame to the panorama.
5. Map the panorama to the rink in the real world by selecting 20 corresponding points manually shown in Fig. 4 and obtain homography $H_{pano,rink}$.
6. Compute homography H_{t,ref^i} mapping the frame at time t to the corresponding reference frame.
7. Obtain homography $H_{t,rink}$ mapping the frame at time t to the rink in the real world $H_{t,rink} = H_{pano,rink} \cdot H_{ref^i,pano} \cdot H_{t,ref^i}$.

---

5. Tracking method

5.1 Hierarchical model

During the hot short track race, the competitors skate rather quickly about less than 10 seconds one loop (110 meters) and the camera pans rapidly in order to capture them. Therefore, the perfect tracking performance seems to be more and more difficult due to the following reasons: (1) The skater's size in the video sequence changes violently. (2) While skating through the top straight on the ice, the skaters are very close to the miscellaneous advertisements attached to the protection board, its color is similar to the one of the skater's clothing sometimes. (3) Occlusions among high speed skaters often happen on the curves of the rink.

All the above challenge the traditional tracking methods, which can't work perfectly. Therefore, to overcome these problems, more contextual knowledge and multiple cues are to be introduced into the tracker as follows: (1) In our application, the short track skater can be considered as a hierarchical model, it consists of two block components: one is the skater's helmet which is a small block, the other bigger one is the body. The relations of their positions are varied due to the skater's diversified postures, when the skater runs through different sub-regions in the rink, such as straight bottom, right curve or left curve, as illustrated in Fig. 5. On considering the camera plan and the skaters' occlusions in the practical application, the helmet block is more discriminative and more reliable than the

body's. Hence, the skater's model including the helmet and the body has been extracted as the prior knowledge, which can effectively reduce the error during the update of the scale of the model. Inaccurate update can cause the tracking to fail when the target model size changes dramatically. (2) Compared with a single-part model, the hierarchical model can reduce the impact of the interference color coming from the advertisement board. At the same time, the integration of the template matching approach (for helmet model) and the color histogram matching method (for body model, details in the section 5.2) is used to solve the problem of occlusions, the benefit of it is to make the tracker more robust under complex environments.
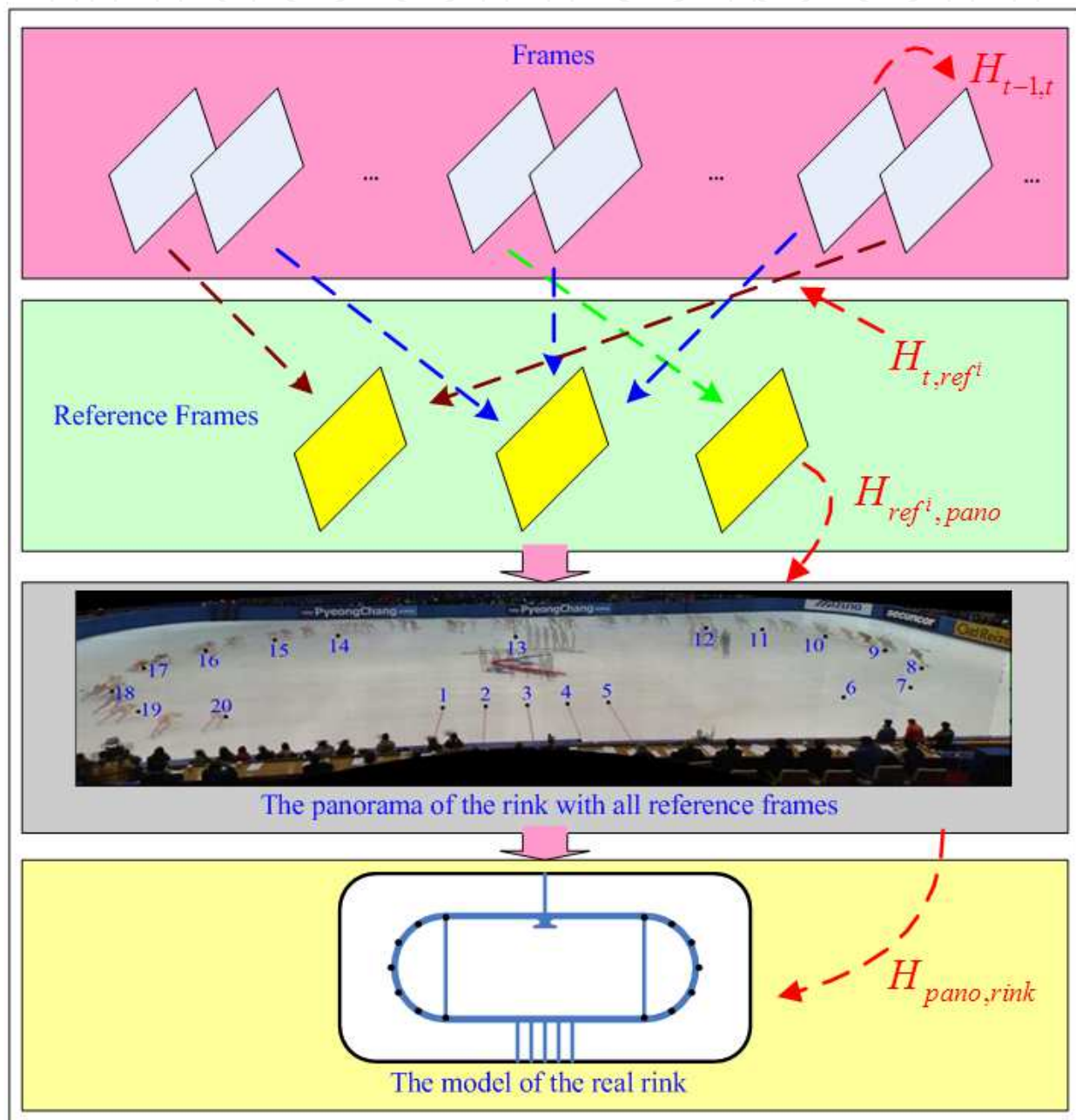

Fig. 4. The flowchart of the registration method for a long image sequence.
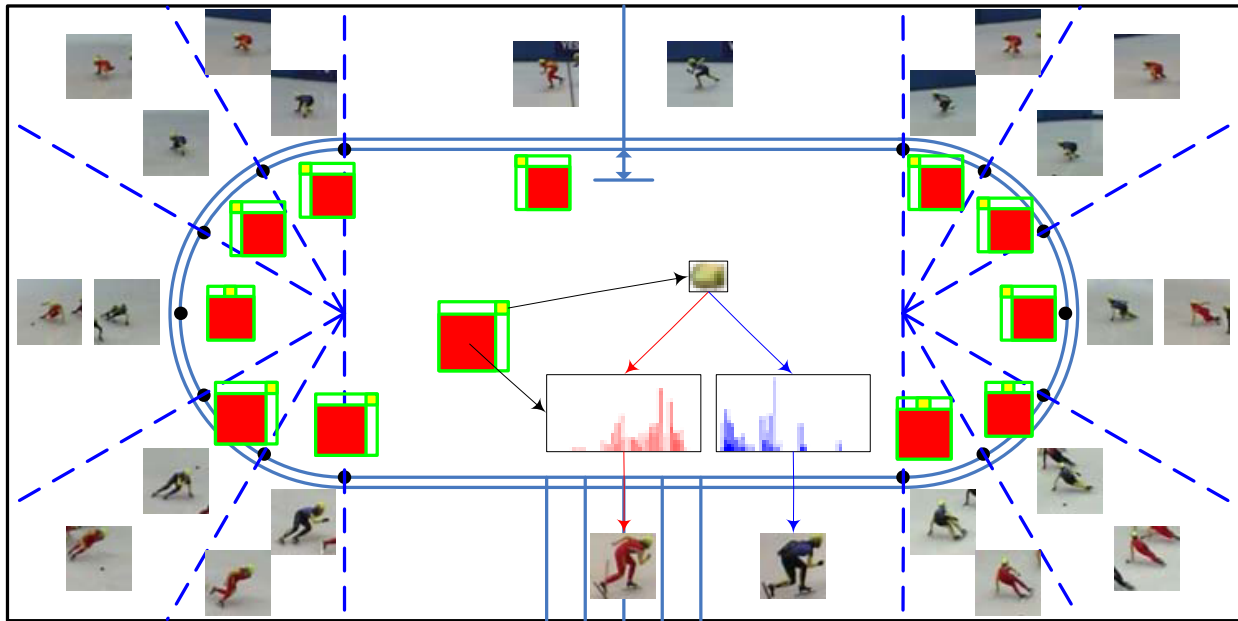
Fig. 5. There are two skaters coming from different countries on the rink. The rink can be divided into 12 sub-regions by the blue dashed lines according to the skater's diversified size and the black markers on the both sides curves. The skater's model of each sub-region differs in its size and the relative position of the helmet and body. The skater's model can be considered as a hierarchical model and represented by two blocks, the smaller yellow block denotes the helmet, the bigger red one denotes the body.

5.2 Color histogram matching method

A skater's body as a target is represented by a rectangle region, its color model can be obtained by using histogram techniques which achieve robustness against non-rigidity, rotation, and partial occlusion (Nummiaro et al., 2003). Let $\{x_i\}_{1,\dots,N}$ be the pixel locations of the target represented by a rectangle region, the region is centered at $\mathbf{z}$ and r is the length of its half diagonal. The function $b(x_i)$ ($\mathbb{R}^2 \rightarrow \{1,\dots,m\}$) assigns the color feature at location x_i to the corresponding bin. To increase the reliability of the color histogram, a weighting function $k(\| x \|^2)$ is employed to assign smaller weights to the boundary pixels farther from the center since those pixels are often affected by occlusions and interference from the background, and denoted as

$$k(\| x \|^2) = \begin{cases} c(1-\| x \|^2) & \| x \|^2 < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where c is a constant, for simple implementation, $c = 1$, if $c = \frac{1}{2} c_d^{-1}(d+2)$, c_d is the volume of the unit d-dimentional sphere, $k(\| x \|^2)$ is the Epanechnikov kernel (Comaniciu et al., 2003). The color histogram $p = \{p_u(z)\}_{u=1\dots m}$ at location z is then calculated as

$$p_u(z) = C \sum_{i=1}^{N} k\left(\left\| \frac{z - x_i}{r} \right\|^2 \right) \delta\left[b(x_i) - u \right] \tag{2}$$

where δ is the Kronecker delta function, C is a normalization constant and derived by imposing the condition $\sum_{u=1}^{m} p_u = 1$.

In our application, the Hue Saturation Value (HSV) color space ($8 \times 8 \times 4$ bins) was used to reduce sensitivity to lighting conditions. To avoid the interference from the background (i.e. the rink), those pixels belonging to the rink are to be kept away from the calculation of the histogram. The function $m(x_i)$ is considered as a mask to filter the rink, and defined as

$$m(x_i) = \begin{cases} 1 & S(x_i) > \tau_s \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $S(x_i)$ is the value of Saturation at location x_i, τ_s is a threshold and $\tau_s = 0.15$ in our experiments. Therefore, the new color histogram is calculated by

$$p_u(z) = C_m \sum_{i=1}^{N} m(x_i) k \left(\left\| \frac{z - x_i}{r} \right\|^2 \right) \delta \left[b(x_i) - u \right] \tag{4}$$

where C_m is a normalization constant.

The Bhattacharyya coefficient (Aherne et al., 1997) as a popular similarity measure between two color histograms $p = \{p_u\}_{u=1...m}$ and $q = \{q_u\}_{u=1...m}$ is denoted by

$$\rho[p,q] = \sum_{u=1}^{m} \sqrt{p_u q_u} \tag{5}$$

The larger ρ is, the more similar the two color histograms. $\rho = 1$ means a perfect match if and only if the two color histograms are identical.

The benefit of our new color histogram is shown in Fig. 6. (a) includes a skater with red clothing, (f) includes the one with blue clothing and is very different from (a), (k) includes the same skater as (a) distracted by another skater with blue clothing, (k) is more similar to (a) than (f) apparently. As a similarity measure, the Bhattacharyya coefficient between (a) and (k) should be larger than the one between (a) and (f). However, calculated the Bhattacharyya coefficients by using the traditional histogram and the weighted histogram, $\rho[p_c, p_m] = 0.677$ and $\rho[p_d, p_n] = 0.689$ are less than $\rho[p_c, p_h] = 0.684$ and $\rho[p_d, p_i] = 0.704$, respectively. Compared with them, our result is that $\rho[p_e, p_o] = 0.769$ is far larger than $\rho[p_e, p_j] = 0.092$, obviously, our result is better.

5.3 The proposed tracking method

The unscented Kalman filter (UKF) was introduced by Julier (Julier et al., 1995; Julier & Uhlmann, 1997) to address the nonlinear state estimation in control theory as a recursive minimum mean square error estimator. Compared with the extended Kalman filter (EKF), the UKF does not approximate the non-linear process and observation model, it uses a set of sigma points to capture the mean and covariance of the system and propagates these sigma points through the dynamic and measurement models without linearization, the UKF is superior to the EKF both in theory and in many practical applications (van der Merwe et al., 2000; Chen et al., 2006).
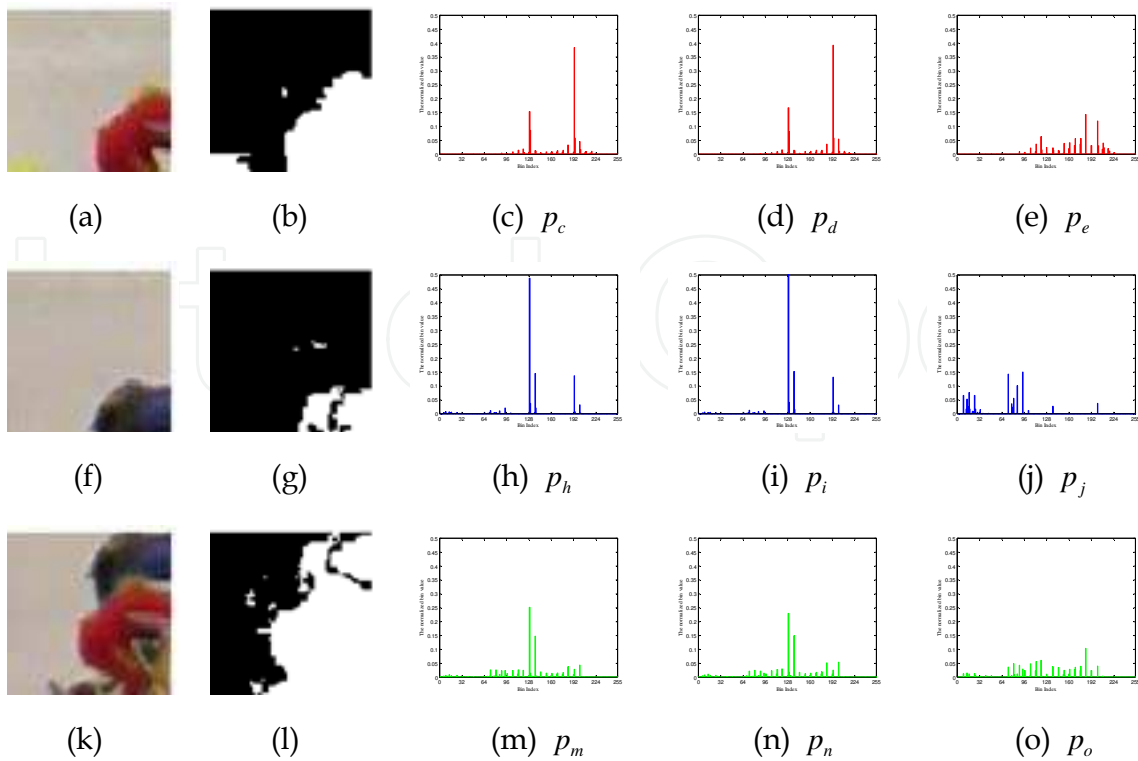
Fig. 6. The benefit of our new color histogram. The first column, (a), (f) and (k) are three different target regions, the second column the results of filtering the rink, the third column the traditional histogram, the fourth column the weighted histogram given by Eq. (2), the last column the new proposed histogram calculated by Eq. (4).

The general state-space model is made up of a state transition $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ and state measurement model $p(\mathbf{y}_t | \mathbf{x}_t)$, and the dynamic model can be denoted as follows:

$$\mathbf{x}_t = \mathrm{f}(\mathbf{x}_{t-1}) + \mathbf{u}_{t-1} \tag{6}$$

$$\mathbf{y}_t = \mathrm{h}(\mathbf{x}_t) + \mathbf{v}_t \tag{7}$$

where $\mathbf{x}_t \in \mathbb{R}^{n_x}$ represents the system state at time t and $\mathbf{y}_t \in \mathbb{R}^{n_y}$ the observation, $\mathbf{u}_t \in \mathbb{R}^{n_u}$ the process noise, and $\mathbf{v}_t \in \mathbb{R}^{n_v}$ the measurement noise. The mapping $\mathrm{f}(\cdot)$ denotes system state model and $\mathrm{h}(\cdot)$ denotes measurement model, respectively.

Here, the proposed tracker combining the UKF with a hierarchical model based on contextual knowledge is as follows:

State $\mathbf{x}_t$ and observation $\mathbf{y}_t$ can be expressed by:

$$\mathbf{x}_t = \left\{ \mathbf{S}_t^h, \mathbf{V}_{t-1} \right\} \tag{8}$$

$$\mathbf{S}_t^h = \left\{ Pos_x^h, Pos_y^h \right\}, \mathbf{V}_{t-1} = \mathbf{S}_{t-1}^h - H_{t-2,t-1} \mathbf{S}_{t-2}^h \tag{9}$$

$$\mathbf{y}_t = \left\{ \mathbf{S}_t^h \right\} \tag{10}$$

In our application, contextual knowledge (CK) as a control input in the observation process is very important for the proposed tracker, it can be consider as a set of internal parameters determined by $\mathbf{S}_t^h$.

$$\mathbf{CK}_t = \left\{ H_{t,rink}, \mathbf{M}_t^h, \mathbf{M}_t^b, PosRel_t^{hb} \right\} \tag{11}$$

$$\mathbf{M}_t^h = \left\{ \mathbf{S}_t^h, R_x^h, R_y^h \right\}, \mathbf{M}_t^b = \left\{ Hist, \mathbf{S}_t^b, R_x^b, R_y^b \right\}, \mathbf{S}_t^b = \left\{ Pos_x^b, Pos_y^b \right\} \tag{12}$$

where at time t, superscript h and b denote helmet and body, respectively, suffix x and y are the coordinates, $\mathbf{V}$ the velocity, R the length of block, Pos the position of the block center, and $PosRel_t^{hb}$ the relation of the helmet relative to the body shown in Fig. 5, and $PosRel_t^{hb} \in \{left, middle, right\}$. $Hist$ is the histogram of the skater's body model. Generally, the helmet model $\mathbf{M}_t^h$ is firstly determined by the template matching method, then $\mathbf{S}_t^b$ is calculated by $\mathbf{S}_t^h$ and $PosRel_t^{hb}$, and $Hist$ can be obtained inside a rectangle region which is computed by $\mathbf{S}_t^b$, R_x^b and R_y^b. Finally, $\mathbf{M}_t^b$ is determined.

Here, the mappings $f(\cdot)$ and $h(\cdot)$ can be expressed by the following equations

$$\mathbf{x}_t = f\left(\mathbf{x}_{t-1}, H_{t-1,t} \right) + \mathbf{u}_{t-1} = \mathbf{A} \left[H_{t-1,t} \left(\mathbf{x}_{t-1} \right) \right] + \mathbf{u}_{t-1} \tag{13}$$

$$\mathbf{y}_t = h\left(\mathbf{x}_t \right) + \mathbf{v}_t = \mathbf{C} \mathbf{x}_t + \mathbf{v}_t \tag{14}$$

where $H_{t-1,t}$ as a control input is used to remove the camera motion expressed by $\left[Pos_x \; Pos_y \; 1 \right]^T = H_{t-1,t} \left[Pos_x^h \; Pos_y^h \; 1 \right]^T$, ($Pos_x^h \in \mathbf{x}_{t-1}$ and $Pos_x^h \in \mathbf{x}_{t-1}$), then the value of Pos_x^h and Pos_x^h are replaced by Pos_x and Pos_x as $Pos_x^h = Pos_x, Pos_y^h = Pos_y$. $\mathbf{A}$ is the state transition matrix, $\mathbf{C}$ is a $n_y \times n_x$ matrix which relates the state to the measurement. The noise term $\mathbf{u}_t$ is a Gaussian random variable with zero mean and a covariance matrix $\mathbf{Q}$, i.e., its probability distribution is $p(\mathbf{u}) \sim N(0, \mathbf{Q})$, the covariance matrix $\mathbf{Q}$ is referred to as the process noise covariance matrix. Much like $\mathbf{u}_t$ for the process, $p(\mathbf{v}) \sim N(0, \mathbf{R})$ and $\mathbf{R}$ is measurement noise covariance matrix.

The UKF tracker can be initialized by:

$$\bar{\mathbf{x}}_0 = E\left[\mathbf{x}_0 \right] \tag{15}$$

$$\mathbf{P}_0 = E\left[\left(\mathbf{x}_0 - \bar{\mathbf{x}}_0 \right) \left(\mathbf{x}_0 - \bar{\mathbf{x}}_0 \right)^T \right] \tag{16}$$

$$\lambda = \alpha^2 \left(n_x + \kappa \right) - n_x \tag{17}$$

$$W_0^{(m)} = \lambda / \left(n_x + \lambda \right) \tag{18}$$

$$W_0^{(c)} = \lambda / \left(n_x + \lambda \right) + \left(1 - \alpha^2 + \beta \right) \tag{19}$$

$$W_i^{(m)} = W_i^{(c)} = 1 / \left\{ 2\left(n_x + \lambda \right) \right\}, i = 1, \ldots, 2n_x \tag{20}$$

where the constant α determines the spread of the sigma points around $\bar{\mathbf{x}}$ and is usually set to a small positive value. The constant κ is a secondary scaling parameter which is usually set to 0 or $3 - n_x$ (Julier et al., 1995). β is a non-negative weighting term which can be used to incorporate knowledge of the higher order moments of the distribution of $\mathbf{x}$, for a Gaussian distribution, and $\beta = 2$ is optimal (van der Merwe et al., 2000). W_i is the weight associated with the *ith* sigma point. Here, these parameters were set to $\alpha = 1$, $\beta = 2$ and $\kappa = 0$.

For $t \in \{1, \dots, \infty\}$, the tracking process combined with UKF and a hierarchical model is detailed as follows:

1. Calculate sigma points

$$\mathcal{X}_{t-1} = \left[\bar{\mathbf{x}}_{t-1} \quad \bar{\mathbf{x}}_{t-1} \pm \sqrt{(n_x + \lambda) \mathbf{P}_{t-1}} \right] \tag{21}$$

where $\sqrt{(n_x + \lambda) \mathbf{P}_{t-1}}$ is the *ith* row or column of the matrix square root of $(n_x + \lambda) \mathbf{P}_{t-1}$ which can be implemented directly by using a Cholesky factorization.

2. Time update:

$$\mathcal{X}_{t|t-1} = \mathrm{f}\left(\mathcal{X}_{t-1}, H_{t-1,t} \right) \tag{22}$$

$$\bar{\mathbf{x}}_{t|t-1} = \sum_{i=0}^{2n_x} W_i^{(m)} \mathcal{X}_{i,t|t-1} \tag{23}$$

$$\tilde{\mathbf{x}}_{i,t|t-1} = \mathcal{X}_{i,t|t-1} - \bar{\mathbf{x}}_{t|t-1} \tag{24}$$

$$\mathbf{P}_{t|t-1} = \sum_{i=0}^{2n_x} W_i^{(c)} \left(\tilde{\mathbf{x}}_{i,t|t-1} \right) \left(\tilde{\mathbf{x}}_{i,t|t-1} \right)^{\mathrm{T}} + \mathbf{Q} \tag{25}$$

$$\mathcal{Y}_{t|t-1} = \mathrm{h}\left(\mathcal{X}_{t|t-1} \right) \tag{26}$$

$$\bar{\mathbf{y}}_{t|t-1} = \sum_{i=0}^{2n_x} W_i^{(m)} \mathcal{Y}_{i,t|t-1} \tag{27}$$

3. Observe in the search regions determined by sigma points and obtain $\mathbf{y}_t$ as a solution of the maximum a posteriori (MAP) based on its contextual knowledge $\mathbf{CK}_t$:

$$\mathbf{y}_t = \underset{i}{\mathrm{argmax}} \, P\left(\mathbf{y}_{i,t} \middle| \mathbf{CK}_t \right) \tag{28}$$

$$P\left(\mathbf{y}_{i,t} \middle| \mathbf{CK}_t \right) = P\left(M_{i,t}^b \middle| M_{i,t}^h, \mathbf{CK}_t \right) P\left(M_{i,t}^h \middle| \mathbf{CK}_t \right) \tag{29}$$

a. Calculate the observation probability of the helmet model in the search region $P\left(M_{i,t}^h \middle| \mathbf{CK}_t \right)$ which is the score of the template matching method, and map the skater's position to the rink in the world $Pos_{i,rink}$ by multiplying $H_{t,rink}$.

b. Compute which sub-region the skater skates through according to $Pos_{i,rink}$, and obtain the index $iRegion \in \{1,\ldots,12\}$, detailed in Fig. 5.

c. Obtain $PosRel_{i,t}^{hb}$ according to $iRegion$, and the body model $M_{i,t}^{b}$ is determined by $\{M_{i,t}^{h}, PosRel_{i,t}^{hb}\}$.

d. Calculate the Bhattacharyya coefficient ρ by Eq. (5) between the target histogram $Hist \in M_{t-1}^{b}$ and the observation one $Hist \in M_{i,t}^{b}$, and obtain the probability $P\left(M_{i,t}^{b} \middle| M_{i,t}^{h}, \mathbf{CK}_t\right) = \rho$.

4. Measurement update:

$$\tilde{\mathbf{y}}_{i,t|t-1} = \mathcal{Y}_{i,t|t-1} - \overline{\mathbf{y}}_{t|t-1} \tag{30}$$

$$\mathbf{P}_{\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t} = \sum_{i=0}^{2n_x} W_i^{(c)} \left(\tilde{\mathbf{y}}_{i,t|t-1}\right)\left(\tilde{\mathbf{y}}_{i,t|t-1}\right)^{\mathrm{T}} + \mathbf{R} \tag{31}$$

$$\mathbf{P}_{\tilde{\mathbf{x}}_t \tilde{\mathbf{y}}_t} = \sum_{i=0}^{2n_x} W_i^{(c)} \left(\tilde{\mathbf{x}}_{i,t|t-1}\right)\left(\tilde{\mathbf{y}}_{i,t|t-1}\right)^{\mathrm{T}} \tag{32}$$

$$\mathbf{K}_t = \mathbf{P}_{\tilde{\mathbf{x}}_t \tilde{\mathbf{y}}_t} \mathbf{P}_{\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t}^{-1} \tag{33}$$

$$\overline{\mathbf{x}}_t = \overline{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \left(\mathbf{y}_t - \overline{\mathbf{y}}_{t|t-1}\right) \tag{34}$$

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{P}_{\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t} \mathbf{K}_t^{\mathrm{T}} \tag{35}$$

6. Experiments and results

6.1 The output of the proposed system

The trajectory of a skater in an individual race of 500 meters is illustrated in Fig. 7, different color line denotes the trajectory of the skater when he/she skates along different loop, and the order is red, green, blue, black and yellow respectively. In Fig. 8, the velocity of one skater in a individual race of 500 meters is shown, the green line represents the velocity obtained manually as the ground truth, and the black one denotes the result of our tracking algorithm. More information and statistic of the competitions are available such as the trajectory and velocity, which can be further processed and analyzed by the sports experts.

6.2 Tracking results

First, our tracking method compares with the MeanSHIFT and CAMSHIFT algorithm from Open Computer Vision Library (OpenCV, http://sourceforge.net/projects/ opencvlibrary/). The results are illustrated in Fig. 9, green box represents the hierarchical model in the corresponding sub-region. If the target color does not change, the MeanSHIFT and CAMSHIFT trackers are quite robust, but they are easily distracted when similar colors appear in the background, however, our tracker can work well by means of a hierarchical model.
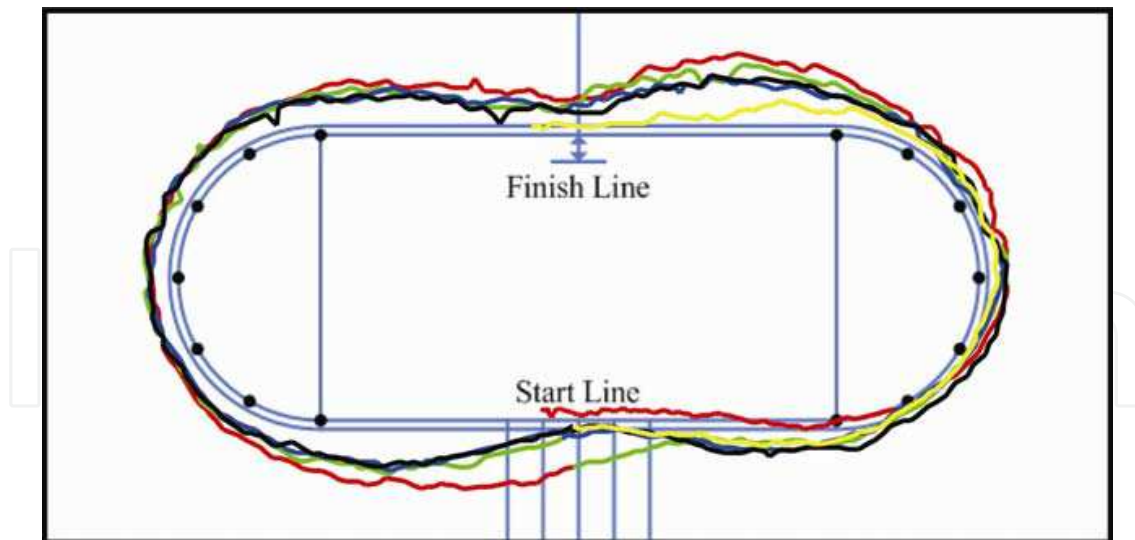
Fig. 7. The trajectory of a skater in an individual race of 500 meters, different color line denotes the trajectory of the skater when he/she skates along different loop.
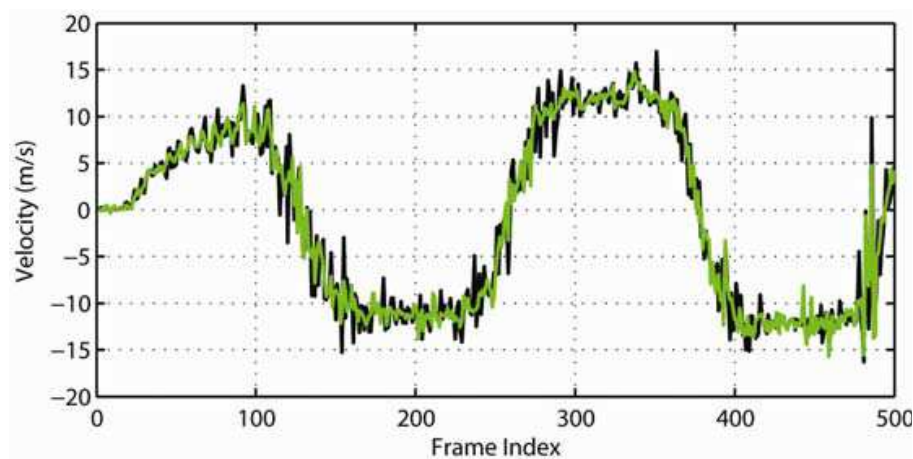

Fig. 8. The velocity of one skater in the individual race of 500 meters.

Second, the comparison results of our tracker and the adaptive color-based particle filter (Nummiaro et al., 2003) are shown in Fig. 10. The failure of the adaptive color-based particle filter tracking is caused by: (1). The model is updated every frame, the patches of the background could be integrated into the model, for a certain time period, the tracker will shift with the background and will not work anymore. (2). Using the single part model (one block or ellipse) and the single cue (color distribution), however, it is too weak for the tracking in the sports domain.

Finally, in Fig. 11, the competitor skating through the curve can be exactly tracked despite successive occlusion. By using the hierarchical model, the proposed tracker would firstly detect the helmet model of the tracked skater by the template matching method. Since the helmet is not occluded, the correct detection result of the helmet model should be observed in the specific search region, then, the color histogram matching method is used to detect the occluded body, compared with other observation results, the observation probability as in Eq. (29) in the specific search region must be the largest, that is, the skater is tracked successfully.
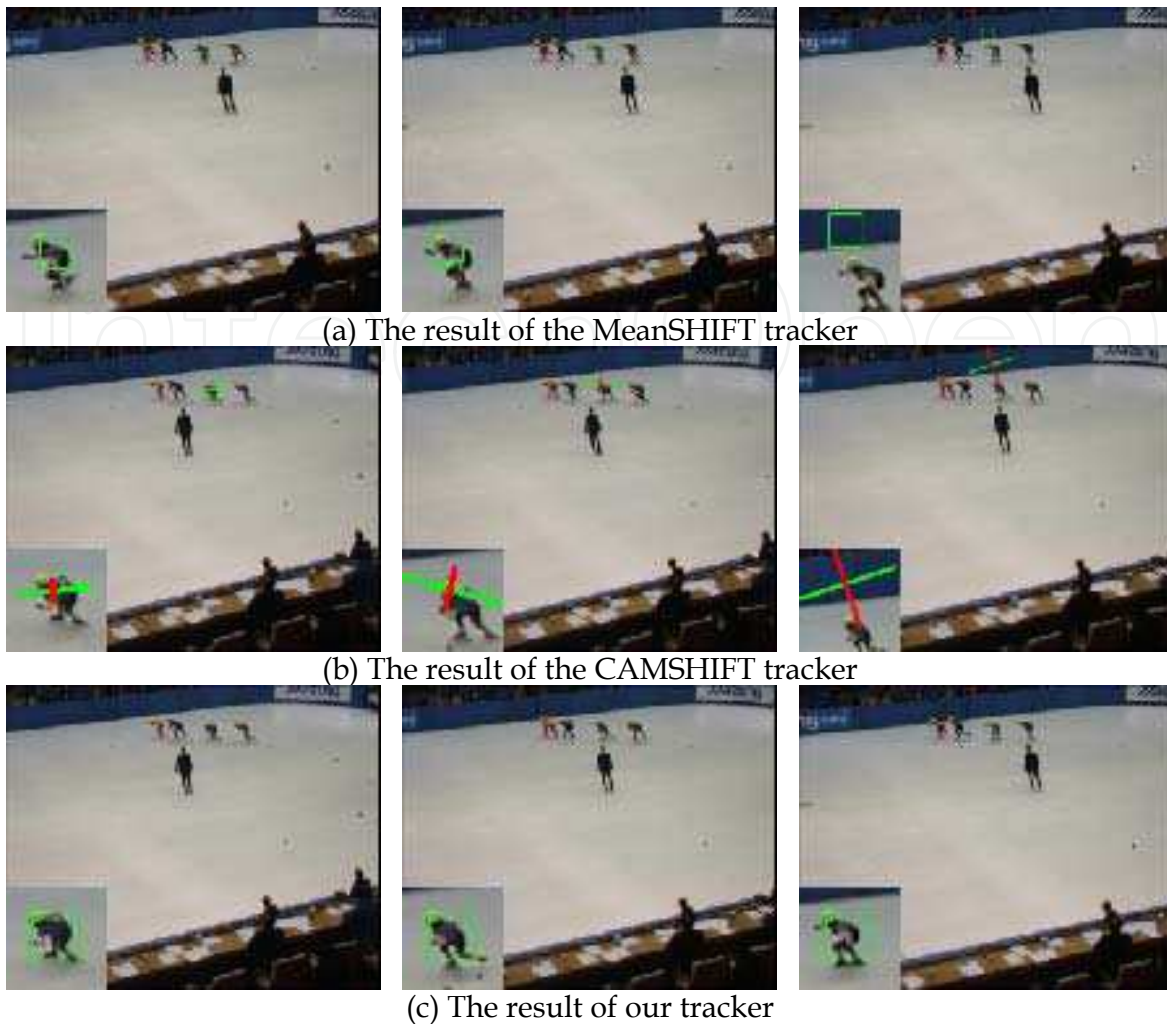
(a) The result of the MeanSHIFT tracker

(b) The result of the CAMSHIFT tracker

(c) The result of our tracker

Fig. 9. MeanSHIFT and CAMSHIFT versus our method.

(a) The result of the adaptive color-based particle filter tracker

(b) The result of our tracker
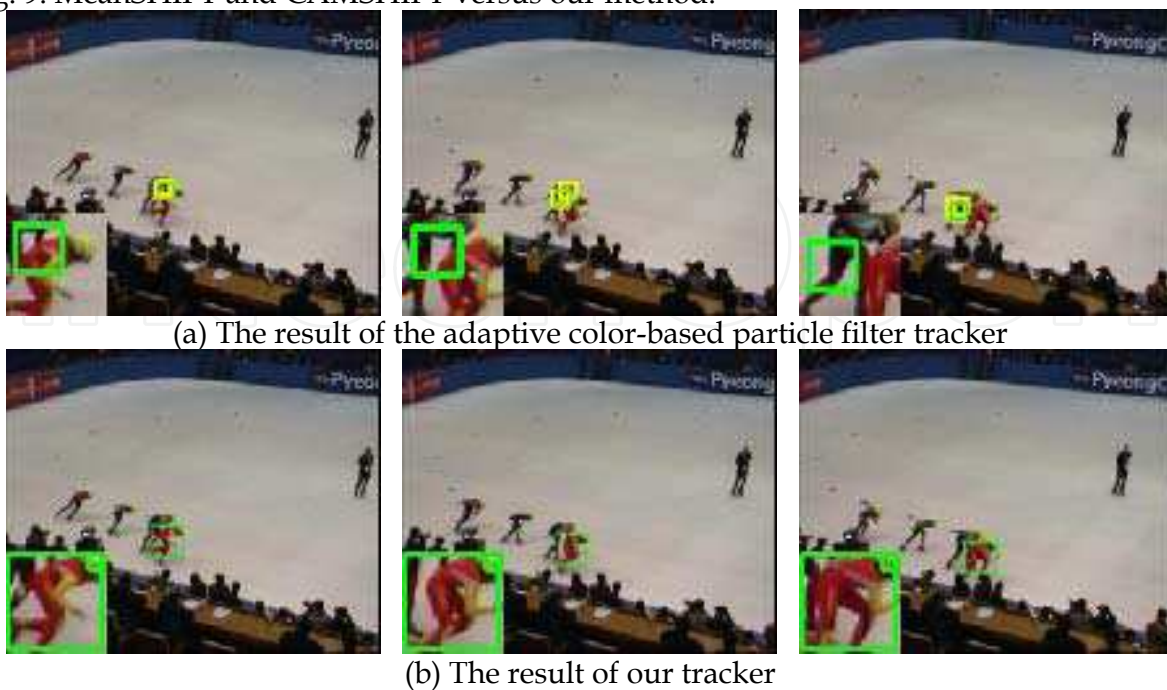
Fig. 10. The adaptive color-based particle filter versus our method.

(a) frame 354   (b) frame 358   (c) frame 362

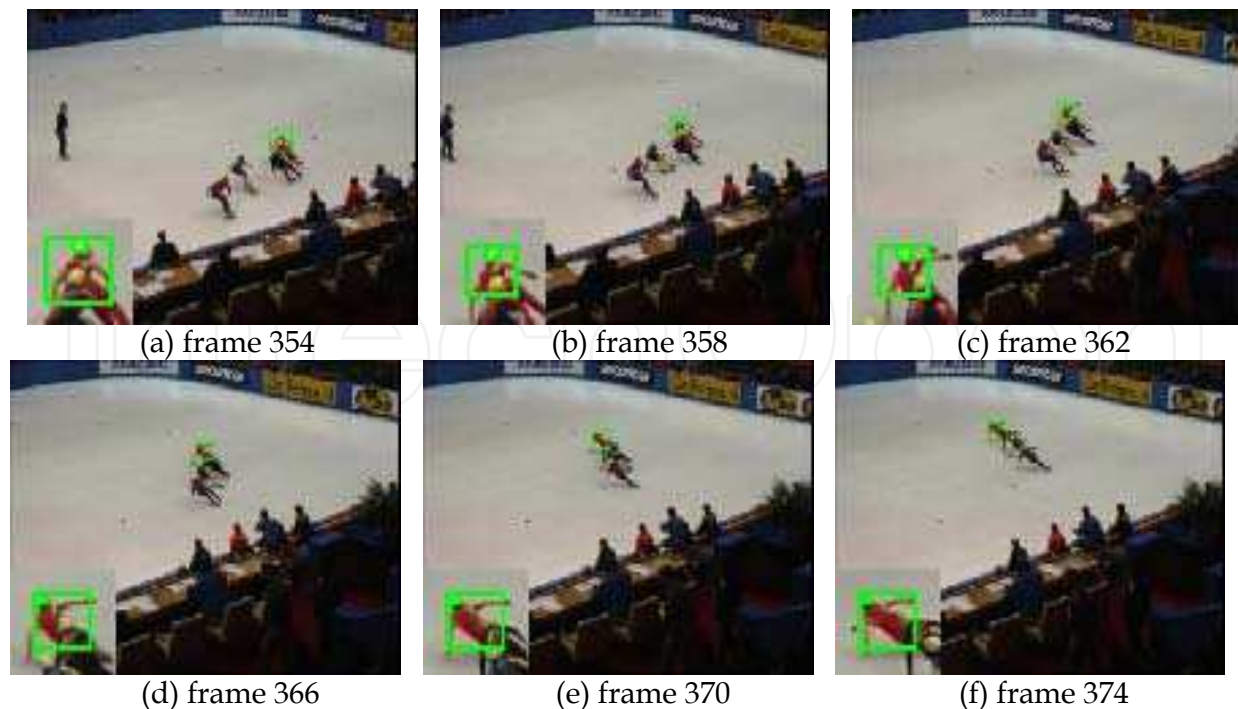(d) frame 366   (e) frame 370   (f) frame 374

Fig. 11. Tracking results of successive occlusion.

The better performance of the proposed tracker is due to the following three aspects:
1. Using the contextual knowledge, namely, the hierarchical model in each sub-region, it can help the tracker to know when and how to update the hierarchical model.
2. Multiple cues: the template matching method (for helmet) and the color histogram matching (for body), it can make the tracker more robust when a skater moving through the advertisement board or occlusions appear on the curve.
3. The unscented Kalman filer can capture the posterior mean and covariance accurately to the 2*nd* order. In addition, compared with the particle filter, the UKF is more efficient since the number of the sigma points increases linearly with the number of dimension.

6.3 Evaluate the homographies

In our practical system as shown in Fig. 2, there are two kinds of homographies: (1) the homography between two neighborhood frames can remove the camera motion to improve the precision of the motion prediction in the process of tracking. (2) Another homography transforming the current frame to the model of the real rink is associated with not only the output of tracking subsystem but also the tracking process to obtain a skater's hierarchical model when the skater moves through different sub-regions of the rink. The calculation of them is a random process and their accuracy is uncertain, therefore, we have to try to evaluate them and analyze how they affect the tracking performance and the system's output.

Note that our camera pans always at the horizontal direction, in which the accuracy of the homographies is very different from that in the vertical direction. Therefore, for all experiments in this section, statistics of variables are to be derived from the two directions respectively. For simplicity's sake, we only give those equations in the horizontal direction which are the same as that in the vertical direction. $|\cdot|_x$ denotes the value of the horizontal direction and $|\cdot|_y$ the value of the vertical direction.

6.3.1 The relation between the accuracy of homography $H_{t-1,t}$ and the tracking performance

	lap 1	lap 2	lap 3	lap 4	lap 5	nframes
BS	1-87	291-348	523-581	756-816	993-1053	326
RC	88-159	349-405	582-636	817-872	1054-1108	295
TS	160-230	406-467	637-700	873-935	1109-1149	301
LC	231-290	468-522	701-755	936-992		227

Table 1. A test video sequence is divided into different groups.

Generally, the tracking process can be considered as two components: *prediction* and *detection* (or observation). Their accuracy can affect the tracking performance directly, however, which one is the main factor leading to the tracking failing: *prediction* or *detection*? It is very important for a practical tracking system. We try to analyze it and test on a video sequence, which are segmented into four groups: bottom straight(BS), top straight(TS), left curve(LC) and right curve(RC), according to the spatio-temporal relation shown in Table 1.
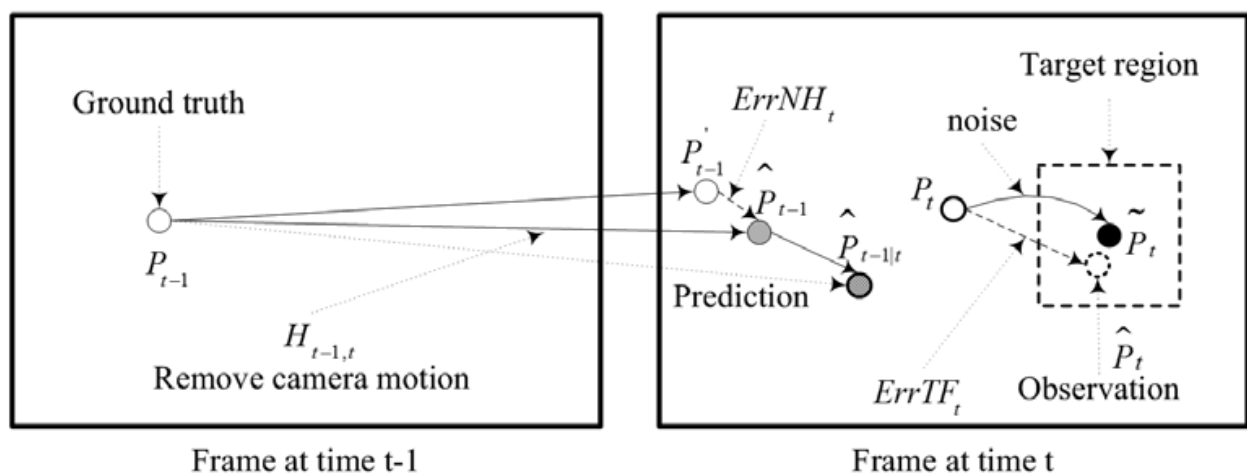

Fig. 12. The diagram of the tracking process between the two neighborhood frames. The ground truth P_t added on Gaussian white noise, i.e. $\tilde{P}_t = P_t + noise$, can be considered as the simulation of the different prediction result $\hat{P}_{t-1|t}$, which assumption is helpful to analyze the tracking performance.

To explain and analyze the tracking process better, some points in the two neighborhood frames are defined, as illustrated in Fig. 12. P_{t-1} is the position of the tracked skater in the frame at time $t-1$ and P_t at time t , which are obtained manually as the ground truth. P_{t-1}' is the corresponding point of P_{t-1} under the ideal transformation if and only if $H_{t-1,t}$ is very accurate, $\hat{P}_{t-1}$ as the approximation of P_{t-1}' is calculated by $\hat{P}_{t-1} = H_{t-1,t}P_{t-1}$. $ErrNH_t$ (the Error of the Neighborhood Homography at time t) is denoted as

$$ErrNH_t = \hat{P}_{t-1} - P_{t-1}' \tag{36}$$

$\hat{P}_{t-1}$ is transformed to $\hat{P}_{t-1|t}$ by using the dynamic model, $\tilde{P}_t = P_t + noise$ can be considered as the simulation of the different prediction result $\hat{P}_{t-1|t}$, which assumption is helpful to

analyze the tracking performance. $ErrTF_t$ (the Error of the Tracking result in the Frame at time t) is expressed by

$$ErrTF_t = \hat{P}_t - P_t \qquad (37)$$

The tracking result would be considered as the accuracy in the ith frame if the distance of $ErrTF_i$ is less than the threshold value τ ($\tau = 5$ pixels) defined as

$$TA_i = \begin{cases} 1 & \text{if } \left\| ErrTF_i \right\|_x \leq \tau \text{ and } \left\| ErrTF_i \right\|_y \leq \tau \\ 0 & \text{otherwise} \end{cases} \qquad (38)$$

where $\left\| \cdot \right\|_x$ denotes the absolute value of the horizontal distance and $\left\| \cdot \right\|_y$ the absolute value of the vertical distance.

As for the image sequence, the precision of the tracking result can be expressed by the following equation, where $nframes$ is the total number of frames used in statistics.

$$TP = \frac{\sum_{i=1}^{nframes} TA_i}{nframes} \qquad (39)$$

	Noise range					
	0	1-5	1-10	5-10	1-15	10-15
BS	100	100	97.55	95.4	88.04	65.34
RC	100	98.99	95.61	90.88	91.99	67.91
TS	100	99.34	98.01	95.35	93.36	83.72
LC	100	100	93.39	86.34	85.46	66.52
All	100	99.57	96.35	92.43	89.9	71.02

Table 2. The precision percentage of the tracking results given six groups of Gaussian white noise.

The experiments have been implemented with six groups of Gaussian white noise. For a better simulation and qualitative analysis, the noise within some range is selected and others are discarded. The noise range is shown in Table 2, $noise = 0$ means that no noise would be added on the ground truth P_t, i.e. the skater's hierarchical model is detected directly on P_t. In Table 2, when $noise = 0$, the precision of the whole match is 100%, that is to say, if the prediction accuracy is enough, our detection method is very robust. However, with the increase of the $noise$ value, i.e, the prediction accuracy becomes much lower, the value of TP decreases gradually. When $noise < 10$, TP in RC and LC is lower than the one in BS and TS due to the skater's occlusion often occurred at RC and LC which can cause our tracker drift or disabled. When $noise > 10$, TP in BS is much lower. In BS, a skater is closer to the camera, its scale changes quickly, its size and the part of the occlusion become bigger, all of which can make the tracker drift out of the threshold value τ easily. At the same time, with the increase of the $noise$ value, the horizontal error of the tracking result in the frame

($| ErrTF_i |_x$) becomes larger in Fig. 13, and the *RMS* (Root Mean Square) error of *ErrTR* (the Error of the Tracking result in the real Rink) defined as Eq. (40) and Eq. (41) is also increasing in Fig. 14.

$$| ErrTR_i |_x = | H_{i,rink} \hat{P}_t - H_{i,rink} P_t |_x = | H_{i,rink} ErrTF_i |_x \tag{40}$$

$$| RMSErrTR |_x = \sqrt{\frac{\sum\limits_{i=1}^{nframes} (| ErrTR_i |_x)^2}{nframes}} \tag{41}$$

All the above demonstrate that the higher the prediction precision is, the better the tracking performance is. However, the prediction result in practice is $\hat{P}_{t-1|t}$, instead of $\tilde{P}_t$, *how accurate the prediction* $\hat{P}_{t-1|t}$ *is?* In Fig. 12, the prediction error comes from two sources: (1) one is the assumption of the uniform motion in the tracking process (i.e. $\hat{P}_{t-1} \rightarrow \hat{P}_{t-1|t}$), but it can be reasonable and acceptable in the practical systems and (2) the other is the homography $H_{t-1,t}$ used to remove the camera motion (i.e. $P_{t-1} \rightarrow \hat{P}_{t-1}$). Therefore, the prediction accuracy depends on the error of the homography $H_{t-1,t}$ (i.e. $ErrNH_t$). To evaluate the homography $H_{t-1,t}$, we have selected 14 still marker blocks on the rink and recorded the positions of those visible markers in each frame as the ground truth, it means that P_{t-1} and P_{t-1}' are prior to be known. The RMS error of $H_{t-1,t}$ is given by

$$| NH |_x = \sqrt{\frac{\sum\limits_{i=1}^{nframes} (| ErrNH_i |_x)^2}{nframes}} \tag{42}$$
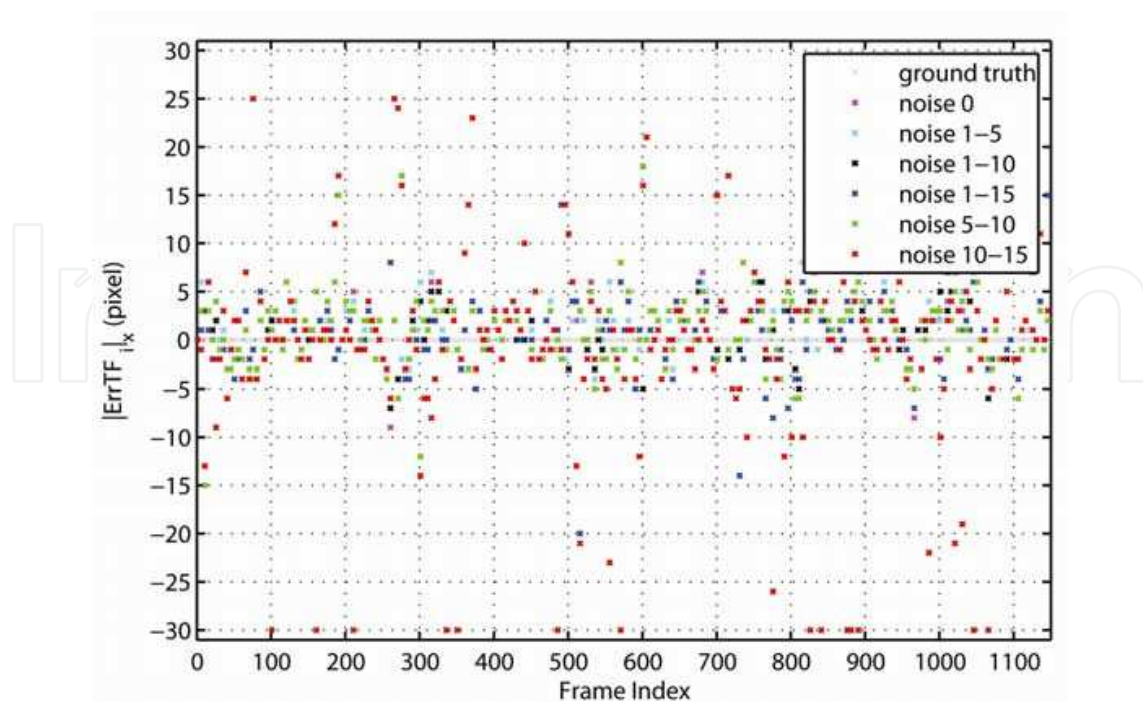

Fig. 13. The horizontal error of the tracking results in the frame given six groups of Gaussian white noise.
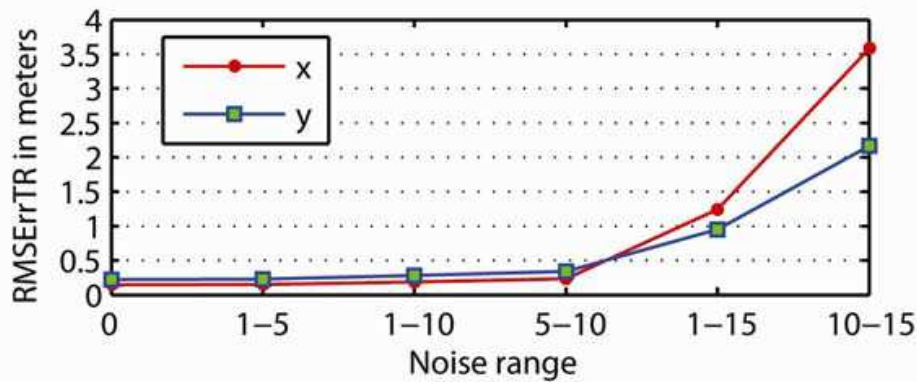
Fig. 14. The RMS error of the tracking result in the real rink given six groups of Gaussian white noise.

	LC							RC						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
$|NH|x$	0.86	1.64	1.28	1.12	1.11	0.99	0.82	1.96	1.92	1.59	1.44	1.15	0.97	1.2
$|NH|y$	0.85	1.21	1.09	1.14	1.13	1.07	0.89	1.29	1.27	1.03	0.95	0.8	0.76	0.81
nframes	181	267	238	187	128	154	231	353	370	381	310	233	248	322

Table 3. The RMS error of $H_{t-1,t}$ of 14 markers on the left curve(LC) and right curve(RC) respectively in pixels.

In Table 3, the accuracy of $H_{t-1,t}$ is enough for the practical application. After all, the calculation of the homography is a random process, but we can increase the number of *good* features (i.e. inliers) to improve the computational stability and reliability of the homograpy (Liu et al., 2007). In a word, the better tracking performance depends on both the robust detection method and the accurate homography $H_{t-1,t}$.

6.3.2 Evaluate the accuracy of homography $H_{i,rink}$ and the precision of our system

The output of our system is the skater's position in the real rink, how to measure the output precision of the practical system needs to be solved. The easiest way is calculating the error between the tracked skater's position and the ground truth (3D position) in the real world directly. However, by this means, we can only obtain the approximation of the former (i.e. $H_{i,rink}P_t$) and can not get the value of the latter without the assistance of the sensors in practice. Therefore, for a more objective evaluation, an indirect approach is used to measure the system's precision (i.e. the accuracy of homography $H_{i,rink}$). In a similar way detailed in section 6.3.1, there are 14 still marker blocks on the rink, their 3D spatial positions are prior to be known according to ISU (http://www.isu.org/) and denoted as RM^j ($j \in \{1,...,14\}$).

The imagery positions of all visible markers in each frame can be recorded manually and transformed to the real rink by the following equation:

$$\widehat{RM}_i^j = H_{i,rink}FM_i^j \tag{43}$$

where FM_i^j is the position of the jth marker in the ith frame, and $\widehat{RM}_i^j$ the estimated position of FM_i^j in the real rink. If $H_{i,rink}$ is always quite accurate, for $\forall i$, $\widehat{RM}_i^j \equiv RM^j$. However, in the practical system, the accuracy of $H_{i,rink}$ is affected by a variety of factors. As the measurement of the registration error caused by $H_{i,rink}$, the standard deviation of $\widehat{RM}_i^j$ and the RMS error are expressed by the following equations:

$$| std^j |_x = \sqrt{\frac{\sum_{i=1}^{nframes}\left(| \widehat{RM}_i^j |_x - E\left[| \widehat{RM}_i^j |_x\right]\right)^2}{nframes}} \tag{44}$$

where $E[\cdot]$ is the expectation function.

$$| rms^j |_x = \sqrt{\frac{\sum_{i=1}^{nframes}\left(| \widehat{RM}_i^j |_x - | RM^j |_x\right)^2}{nframes}} = \sqrt{\frac{\sum_{i=1}^{nframes}\left(| H_{pano,rink}H_{i,pano}FM_i^j |_x - | RM^j |_x\right)^2}{nframes}} \tag{45}$$

where $H_{pano,rink}$ maps the panorama to the real rink, however, $H_{pano,rink}$ is inaccurate in practice, we define the error of $H_{pano,rink}$ as

$$| PRH^j |_x = | H_{pano,rink}PM^j |_x - | RM^j |_x \tag{46}$$

where PM^j is the position of the jth marker based on the panorama, it is recorded manually as the ground truth (see details in Algorithm 1 step 5).

The registration error of 14 markers is shown in Table 4. The values of rms distribute differently, ones are highly accurate and others have large symmetrical errors, all that are derived from the inaccuracy of $H_{pano,rink}$ (i.e. PRH, detailed in last two rows) directly and mostly, since $H_{pano,rink}$ dose not depend on the subscript i in Eq. (45). The value of std (mainly caused by $H_{i,pano}$) is uniformly small, and it means that $H_{i,pano}$ is more accurate than $H_{pano,rink}$. As a result, the accurate $H_{i,rink}$ is due to the better $H_{i,pano}$ and $H_{pano,rink}$. These markers in the rink are closer to the trajectory of a skater, so it is reasonable that the registration error can be considered as an approximation of the output precision of our system. The smaller the values of rms and std are, the more reliable the output of our proposed system is.

	LC							RC						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
|rms|x	0.46	0.31	0.07	0.19	0.16	0.06	0.26	0.43	0.31	0.05	0.26	0.25	0.04	0.29
|rms|y	0.45	0.28	0.14	0.11	0.21	0.21	0.07	0.39	0.17	0.07	0.05	0.11	0.19	0.16
|std|x	0.04	0.04	0.06	0.05	0.04	0.04	0.03	0.05	0.06	0.04	0.04	0.04	0.03	0.02
|std|y	0.06	0.08	0.07	0.06	0.04	0.04	0.03	0.09	0.09	0.07	0.06	0.05	0.03	0.04
nframes	186	280	248	201	139	168	246	365	382	397	326	244	263	339
|PRH|x	0.4	0.21	-0.12	-0.3	-0.18	0.04	0.23	-0.33	-0.27	0.08	0.28	0.31	0.04	0.27
|PRH|y	-0.19	-0.01	0.14	0.13	-0.13	-0.15	0.03	-0.16	0.01	0.15	0.1	-0.02	-0.16	0.08

Table 4. The registration error of 14 markers on the left curve(LC) and right curve(RC) respectively in meters.

7. Conclusion and discussion

In this paper, we propose a novel computer vision system for tracking high-speed non-rigid skaters over a large playing area in short track speed skating competitions. Several important features distinguish the proposed approach from others:

1. Introducing the reference frames as a transition through which each frame can be mapped to the field model to reduce the error accumulation of the projection, and it's very important for a long video sequence and helpful for improving the precision of the system.

2. Incorporating the hierarchical model based on the contextual knowledge and multiple cues into the unscented Kalman filter to improve the tracking performance when occlusions occur.

3. Choosing the unscented Kalman filter for visual tracking in the sports domain, it is superior to EKF in theory, and is more efficient than particle filter.

4. Evaluating the relation between the accuracy of homography $H_{t-1,t}$ and tracking performance.

5. Proposing a novel and objective evaluation method to measure the precision of our practical system.

However, The main problem is remained in our current system: how to improve the tracking performance when skaters are moving in groups during a long and continuously full occlusion.

In future, we can model the skater's uniform of different teams in each sub-region, and the uniform model can be used to assist in tracking the target occluded for a long time.

8. Acknowledgements

9. References

Aherne, F. J.; Thacker, N. A. & Rockett, P. I. (1997). The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data, Kybernetika, vol. 34, No. 4, pp. 363-368, ISSN: 0023-5954

Brown, M. & Lowe, David G. (2003). Recognising panoramas, Ninth IEEE International Conference on Computer Vision, pp. 1218-1225, ISBN: 0-7695-1950-4, Nice, France, October 2003

Chen, Yunqiang; Rui, Yong & Huang, Thomas S. (2006). Multicue HMM-UKF for Real-Time Contour Tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, No. 9, pp. 1525-1529, ISSN: 0162-8828

Comaniciu, D.; Ramesh, V. & Meer, P. (2003). Kernel-based object tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, No. 5, pp. 564-577, ISSN: 0162-8828

Fischler, M. A. & Bolles, Robert C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Communications of the ACM, vol. 24, No. 6, pp. 381-395, ISSN: 0001-0782

Farin, Dirk; Krabbe, Susanne; de With, Peter H.N. & Effelsberg, Wolfgang (2004). Robust Camera Calibration for Sport Videos Using Court Models, SPIE Storage and

Retrieval Methods and Applications for Multimedia, pp. 80-91, ISBN: 0-8194-5210-6, San Jose CA, January 2004

Haritaoglu, I.; Harwood, D. & Davis, L. S. (2000). W4: Real-Time Surveillance of People and Their Activities, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, No. 8, pp. 809-830, ISSN: 0162-8828

Harris, Chris & Stephens, Mike (1988). A combined corner and edge detector, Fourth Alvey Vision Conference, pp. 147-151, University of Manchester, September, 1988

Hartley, R. & Zisserman, A. (2000). Multiple View Geometry in Computer Vision, Cambridge University Press, ISBN: 0-5216-2304-9

Hayet, Jean-Bernard; Piater, Justus & Verly, Jacques (2004). Robust incremental rectification of sport video sequences, British Machine Vision Conference, Kingston University, London, September, 2004

Hayman, Eric & Murray, David W. (2003). The Effects of Translational Misalignment when Self-Calibrating Rotating and Zooming Cameras, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, No. 8, pp. 1015-1020, ISSN: 0162-8828

Intille, S. S. & Bobick, A. F. (1994). Tracking Using a Local Closed-World Assumption: Tracking in the Football Domain, MIT Media Lab Perceptual Computing Group Technical Report 296, 1994

Jug, M.; Pers, J.; Dezman, B. & Kovacic, S. (2003). Trajectory Based Assessment of Coordinated Human Activity, Third International Conference ICVS 2003, pp. 534-543, ISBN: 3-540-00921-3, Graz, Austria, April 2003

Julier, Simon J. & Uhlmann, Jeffrey K. (1997). A New Extention of the Kalman Filter to Nonlinear Systems, Proceedings of the 11th Annual International Symposium on Aerospace/Defense Sensing, Simulation, and Controls, Orlando, Florida, April 1997

Julier, Simon J.; Uhlmann, Jeffrey K. & Durrant-Whyte, Hugh F. (1995). A new approach for filtering nonlinear systems, Proceedings of the American Control Conference, pp. 1628-1632, ISBN: 0-7803-2446-3, Seattle, Washington, June 1995

Junior, B. M. & Anido, R. D. O. (2004). Distributed real-time soccer tracking, Proceedings of the ACM Second International Workshop on Video Surveillance & Sensor Networks, pp. 97-103, ISBN: 1-58113-934-9, New York, October 2004

Liu, GuoJun; Tang, XiangLong; Sun, Da & Huang, JianHua (2007). Robust Registration of Long Sport Video Sequence, Proceedings of the 5th International Conference on Computer Vision Systems, ISBN: 3-0002-0933-8, Bielefeld University, March 2007

Lowe, David G. (2004). Distinctive image features from scale-invariant keypoints International Journal of Computer Vision, vol. 60, No. 2, pp. 91-110, ISSN: 1573-1405

Misu, Toshihiko; Naemura, Masahide; Zheng, Wentao; Izumi, Yoshinori & Fukui, Kazuo (2002). Robust tracking of soccer players based on data fusion, 16th International Conference on Pattern Recognition, pp. 556-561, ISBN: 0-7695-1695-X, Quebec, Canada, August 2002

Needham, C. J. & Boyel, R. D. (2001). Tracking multiple sports players through occlusion, congestion and scale, Proceedings of the British Machine Vision Conference, pp. 93-102, ISBN: 1-901725-16-2, Manchester, UK, September 2001

Nummiaro, K.; Koller-Meier, E. & Gool, L. V. (2003). An adaptive color-based particle filter Image and Vision Computing, vol. 21, No., pp. 99-110, ISSN: 0262-8856

Okuma, K. (2003). Automatic acquisition of motion trajectories, Master's thesis, University of British Columbia, 2003

Okuma, Kenji; Little, James J. & Lowe, David G. (2004). Automatic rectification of long image sequences, Asian Conference on Computer Vision, 2004

Ozyildiz, E.; Krahnstover, N. & Sharma, R. (2002). Adaptive texture and color segmentation for tracking moving objects Pattern Recognition, vol. 35, No., pp. 2013-2029, ISSN: 0031-3203

Pers, J.; Bon, M.; Kovacic, S.; M. Sibila & Dezman, B. (2002). Observation and Analysis of Large-scale Human Motion Human Movement Science, vol. 21, No. 2, pp. 295-311, ISSN: 0167-9457

Pers, J.; Vuckovic, G.; Kovacic, S. & Dezman, B. (2001). A low-cost real-time tracker of live sport events, International symposium on image and signal processing and analysis, pp. 362-365, June 2001

Pingali, G. S.; Jean, Y. & Carlborn, I. (1998). Real time tracking for enhanced tennis broadcasts, International Conference on Computer Vision and Pattern Recognition, pp. 260-265, ISBN: 0-8186-8497-6, Santa Barbara, CA, June 1998

Rasmussen, Christopher & Hager, Gregory D. (2001). Probabilistic data association methods for tracking complex visual objects, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, No. 6, pp. 560-576, ISSN: 0162-8828

Rigoll, G.; Breit, H. & Wallhoff, F. (2004). Robust tracking of persons in real-world scenarios using a statistical computer vision approach, Image and Vision Computing, vol. 22, No., pp. 571-582, ISSN: 0262-8856

Shi, Jianbo & Tomasi, Carlo (1994). Good Features to Track, International Conference on Computer Vision and Pattern Recognition, Seattle, Washington, June 1994

Shum, Heung-Yeung & Szeliski, Richard (2000). Systems and Experiment Paper: Construction of Panoramic Image Mosaics with Global and Local Alignment, International Journal of Computer Vision, vol. 36, No. 2, pp. 101-130, ISSN: 1573-1405

Sivic, Josef & Zisserman, Andrew (2003). Video Google: A Text Retrieval Approach to Object Matching in Videos, Ninth IEEE International Conference on Computer Vision, pp. 1470-1477, ISBN: 0-7695-1950-4, Nice, France, October 2003

Tomasi, Carlo & Kanade, Takeo (1991). Detection and Tracking of Point Features, Technical Report CMU-CS-91-132, Carnegie Mellon University, April 1991

van der Merwe, Rudolph; Doucet, Arnaud; de Freitas, Nando & Wan, Eric (2000). The Unscented Particle Filter, Technical Report CUED/FINFENG/TR380, Cambridge University Engineering Department, August 2000

Vandenbroucke, N.; Macaire, L. & Postaire, J. G. (2003). Color image segmentation by pixel classification in an adapted hybrid color space: Applications to soccer image analysis, Computer Vision and Image Understanding, vol. 90, No., pp. 190-216, ISSN: 1077-3142

Yan, Fei; Kostin, Alexey; Christmas, William J. & Kittler, Josef (2006). A Novel Data Association Algorithm for Object Tracking in Clutter with Application to Tennis Video Analysis International Conference on Computer Vision and Pattern Recognition, pp. 634-641, ISBN: 0-7695-2597-0, New York, June 2006

Yang, Changjiang; Duraiswami, R. & Davis, L. (2005). Efficient mean-shift tracking via a new similarity measure International Conference on Computer Vision and Pattern Recognition, pp. 176-183, ISBN: 0-7695-2372-2, San Diego, CA, June 2005

Yilmaz, A.; li, Xin & Shah, M. (2004). Contour-based object tracking with occlusion handling in video acquired using mobile cameras, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, No. 11, pp. 1531-1536, ISSN: 0162-8828

Kalman Filter

Edited by Vedran Kordic

The Kalman filter has been successfully employed in diverse areas of study over the last 50 years and the chapters in this book review its recent applications. The editors hope the selected works will be useful to readers, contributing to future developments and improvements of this filtering technique. The aim of this book is to provide an overview of recent developments in Kalman filter theory and their applications in engineering and science. The book is divided into 20 chapters corresponding to recent advances in the filed.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds