Tracking Moving Objects in Road Traffic Sequences

Salma Kammoun Jarraya[1], Najla Bouarada Ghrab[1],
Mohamed Hammami[2], and Hanene Ben-Abdallah[1]

[1] MIRACL-FSEG, Sfax University, Rte Aeroport Km 4, 3018 Sfax, Tunisia
[2] MIRACL-FS, Sfax University, Rte Soukra Km 3, 3018 Sfax, Tunisia
{Salma.Kammoun,Hanene.Benabdallah}@fsegs.rnu.tn,
Najla.Bouarada@yahoo.fr, Mohamed.Hammami@fss.rnu.tn

Abstract. In this paper, we present an algorithm for tracking objects in road traffic sequences which is based on coherent strategy. This strategy relies on two times processing. Firstly, a Short-Term Processing (STP) based on spatial analysis and multilevel region descriptors matching allows identification of objects interactions and particular objects states. Secondly, a Long- Term Processing (LTP) is applied to cope with track management issues. In fact LTP feedbacks objects and their corresponding regions in each frame to update tracked object attributes. In case of merging objects, attributes are obtained using Template matching. An experimental study by quantitative and qualitative evaluations shows that the proposed approach can deal with multiple rigid objects whose sizes vary over time. The obtained results prove that our method can provide an effective and stable road objects tracks.

Keywords: Tracking moving object, foreground segmentation, point descriptors, template matching**.**

1 Introduction

Road traffic monitoring has become a very important research area. Such system is based essentially on tracking road objects. The aim of racking object is to estimate the trajectory of moving objects over time. The information gathered by road objects tracking can help to identify their behavior in the observed scene and allows building statistical information about road traffic.

The purpose of our contributions is to track multiple rigid moving objects (road objects) with different sizes and speeds. Note that moving objects are detected automatically. The proposed method for object tracking takes in consideration the possibly states changes of moving objects and interactions between them. In addition, appearance of a new object and disappearance of existing object are managed automatically.

The reminder of this paper is divided into 4 sections. In Section 2, we describe a brief state of art in object tracking. Section 3 presents our proposed method. Section 4 outlines the results of a quantitative and a qualitative evaluation. Finally, Section 5 recapitulates the presented method and outlines future work.

A. Elmoataz et al. (Eds.): ICISP 2012, LNCS 7340, pp. 406–414, 2012.

2 State of the Art in Object Tracking

Several methods [1]-[4] deal with object tracking; however their accuracy depends on, both, constraints and context of the application. Constraints are related to the tracked object(s) (single or multiple, rigid or non-rigid), to the camera (single or multiple, mobile or fixed) and to the observed scene (indoor and/ or outdoor). Dealing with context, we distinguish different applications as person tracking, road object tracking, ball tracking, etc. In this paper, we focus particularly on road object tracking. In addition to constraints of application context, methods reported in literature differ by theirs manner to represent object. We can classify these methods in two categories of approaches which are (1) Points based approach [*cf.* 1,2] and, (2) Model based approach [*cf.* 3,4] (silhouettes or kernel). In these approaches, tracking strategy relies on matching information provided by points/models over times. Points based methods are fast and can deal with partial occlusions. However, cannot usually cope with complex deformation of nonrigid objects. In model based approach, silhouettes model allows tracking of both nonrigid and rigid objects but their computations is very expensive and lack of generality. Unlike silhouettes model, kernel based model can be obtained without knowledge about object nature or shape but cannot resolve occlusions. Within the context of road traffic, methods aim to track unlimited number of rigid road objects in video stream. Thus, we adopt points based approach.

In literature, several methods [1,2][5]-[8] are based on descriptors points. Among of techniques to compute descriptor points are Harris detector [9], KLT (Kanade-Lucas-Tomasi) detector [10] and SIFT descriptor (Scale Invariant Feature Transform)[11]. Our comparative study between these techniques, shows that descriptors from SIFT are invariant to different invariance criteria (*Translation, Scale Changes, Image Rotation, Illumination changes, Image Locale Deformations, and Affine Transformation*). In addition, from theatrical point of view, SIFT technique can (1) produce a great number of descriptor points, (2) give a local image measurement that is robust to noises and to partial occlusions and (3) give distinctive points.

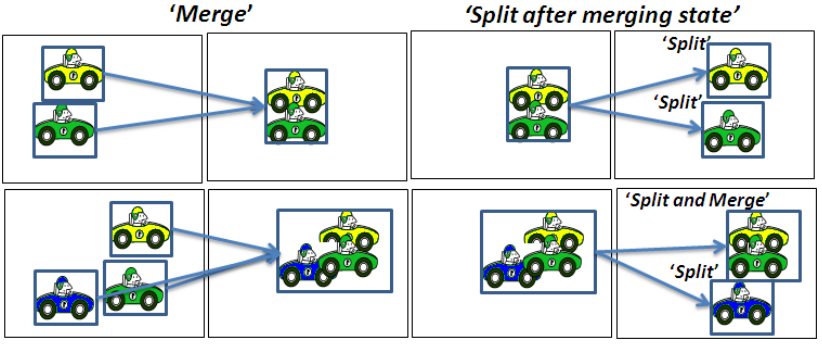

Fig. 1. Examples of interactions between road objects

Note that the success of such tracking field relies on the management of both frequently object state changes (life cycle) and interactions. Life cycle of objects road start by their appearance in the scene (state *'Entry'*) and ended by their disappearance

(state *'Exit'*). During its presence in the scene, a road object can be in normal state (*'Normal'*), normal state with a high speed (*'Normal HS'*), stopped (*'Stopped'*), restart motion after stopped (*'Re-moving'*).

During a life cycle, two types of interactions between objects road can occur (Fig.1). The first interaction happens when two or many objects appear close to each other (*'Merge'*) causing partial or total occlusion. The second interaction results of two or many objects fragmentation (*'Split'*) after merging state.

The most recent tracking methods based on SIFT technique (*cf.* [5]-[8]) track preselected (single or limited number) specific object(s). In addition, states changes of moving objects are not considered. Furthermore, appearance and disappearance of objects are managed according to a region of interest drawing manually.

3 Proposed Method

Our proposed method for tracking road objects is based on three main steps: (1) Foreground segmentation, (2) Short-Term Processing (STP) and (3) Long- Term Processing (LTP). In fact, let $R_t^{cc=1...m}$ and $R_{t-1}^{c=1...n}$ denote respectively the segmented regions from frames F_t and F_{t-1} with $cc \in \{1,...,m\}$ and $c \in \{1,...,n\}$, n and m, are respectively the number of region in F_t and F_{t-1} . Foreground segmentation is done by the method presented in our previous work [12]. This method is based on background modeling approach; it demonstrates robust and accurate results under most of the common problem in foreground segmentation. In STP, $R_t^{cc=1...m}$ and $R_{t-1}^{c=1...n}$ are used to manage objects states and interactions for each input frame, thus produces region correspondence ($R_{j=\{t-1,t\}}^{i=\{c,cc\}}.Cor$) and state ($R_{j=\{t-1,t\}}^{i=\{c,cc\}}.State$). LTP establish all objects tracks ($TrackingObject\{O_{i=1...ObjectCount}\}$) between $t=0$ and t based on $R_{j=\{t-1,t\}}^{i=\{c=1...n,cc=1...m\}}.Cor$ and $R_{j=\{t-1,t\}}^{i=\{c=1...n,cc=1...m\}}.State$ to generate objects trajectories. In the following subsections, we detail the STP and the LTP steps.

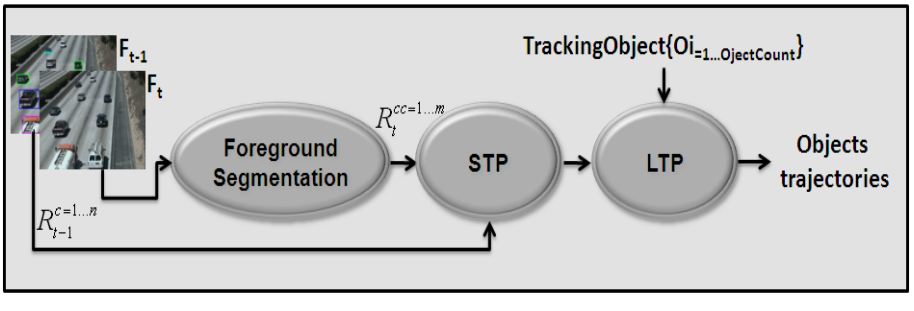

Fig. 2. Flowchart of the proposed tracking method

3.1 Short-Term Processing (STP)

Short-Term Processing takes into account, (1) the spatial analysis and, (2) Multilevel region descriptors matching of $R_t^{cc=1...m}$ and $R_{t-1}^{c=1...n}$. Each region R is represented by a

set of attributes ($Z(R) = (\beta^{1...5}(R), \phi_k^{128}(R))$). Where $\beta^{1...5}(R)$ are 2D spatial attributes (*cf.* Fig. 3) and $\phi_k^{128}(R)$ is a *K-by-128* matrix, each row gives an invariant descriptor for one of the *K* keypoints. The descriptor is a vector of 128 values normalized to unit length. Regions correspondences ($R_{j=\{t-1,t\}}^{i=\{c=1...n,cc=1...m\}}.Cor$) are initialized by -1.

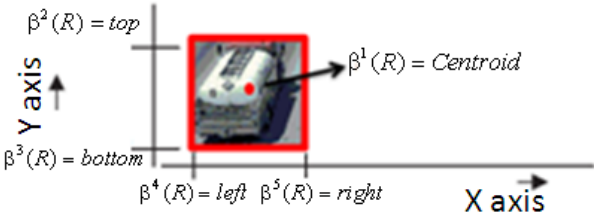

Fig. 3. 2D Spatial attributes ($\beta^{1...5}(R)$)

Spatial Analysis. We project $\beta_t^1(R_t^{cc=1...m})$ onto area from $\beta_t^{2...5}(R_{t-1}^{c=1...n})$, thus provides correspondence for regions in *'Normal'* states and/or in *'Split'* interactions according. Region in state *'Normal'* corresponds to the case where $\beta_t^1(R_t^{cc})$ belongs to only one R_{t-1}^c area. The Split interaction corresponds to the case where β_t^1 of two or more R_t^{cc} $\left(i.e\ R_t^{cc=1,cc=2...}\right)$ belong to one R_{t-1}^c area. We associate regions R_t^{cc} and R_{t-1}^c according to Equation 1.

$$\begin{cases} If\ \ 'Normal'\ state\ then \\ R_t^{cc}.Cor = c \\ Else\ If\ \ 'Split'\ \ interaction\ then \\ R_t^{cc=1,cc=2...}.Cor = c \end{cases} \quad (1)$$

Multilevel Region Descriptors Matching. A multilevel region descriptors matching is proposed for $R_{j=\{t-1,t\}}^{i=\{c=1...n,cc=1...m\}}$ with ($R_{j=\{t-1,t\}}^{i=\{c=1...n,cc=1...m\}}.Cor == -1$). This step allows us to cope with region interaction (*'Merge'*) and states (*'Entry'*, *'Exit'*, *'Normal VE'*, *'Stopped'* and *'Re-Moving'*). We aim to select, for each region descriptors $\phi_{i=1...k_1}^{128}(R_t^{cc})$, its match to $\phi_{k_2}^{128}(R_{t-1}^c)$ (Equation (2)). There is matching (R_Match) between two regions in case of at least one descriptor match (Des_Match). Decision to select matched descriptors from $\phi_k^{128}(R_{t-1}^c)$ is given by Equation (3). In our work, SIFT descriptors matching is based on dot products ($DP^{i=1...k_1}$) between unit vectors of descriptors (Equation (4)). Generic rules of the multilevel region descriptors matching is presented by Algorithm 1.

$$\begin{cases} R\_Match(\phi_{k_1}^{128}(R_t^{cc}), \phi_{k_2}^{128}(R_{t-1}^{c})) = 1 \\ If\ any(Des\_Match > 0) \end{cases} \quad (2)$$

$$\begin{cases} Des\_Match(i)=1 \\ if(DP^{i=1\ldots k_1}(1) < 0.6 * DP^{i=1\ldots k_1}(2) \end{cases} \quad (3)$$

$$DP^{i=1\ldots k_1} = sort(arcosine(\phi_{i=1\ldots k_1}^{128}(R_t^{cc}) * \phi_{k_2}^{128}(R_{t-1}^{c})^T) \quad (4)$$

```
Algorithm 1: Multilevel region descriptors matching

Input: f_{k1}^{128}(R_t^{cc=1...m}), f_{k2}^{128}(R_{t-1}^{c=1...n}), Stopped(O_{j=1...h}).f
Output: R_{j={t-1,t}}^{i={c=1...n,cc=1...m}}.State, R_{j={t-1,t}}^{i={c=1...n,cc=1...m}}.Cor,
Stopped(O_{j=1...h}).f
If R_Match(f_{k1}^{128}(R_t^{cc=1...m}), f_{k2}^{128}(R_{t-1}^{c=1...n})^T) then
  If (f_{k1}^{128}(R_t^{cc}), f_{k2}^{128}(R_{t-1}^{c})^T) then
   R_t^{cc}.Cor = c
  Else If (f_{k1}^{128}(R_t^{cc}), f_{k2}^{128}(R_{t-1}^{c1,c2..cq})^T) then
   R_t^{cc}.Cor = c1,c2..cq
  End
Else
  If R_Match(f_{k1}^{128}(R_t^{cc=1...m}), Stopped(h).f) then
   R_t^{cc}.Cor = Stopped(h).Cor
  Else
   R_t^{cc}.Cor = *
  End
  If R_Match(f_{k2}^{128}(R_{t-1}^{c=1...n}), f_{k3}^{128}(R_t^{c=1...m})) then
   Stopped(j+1).Cor = c
   Stopped(j+1).f = f_{k3}^{128}(R_t^{c})
  Else
   R_{t-1}^{c}.Cor = *
  End
End
```

Three level matching levels are proposed: the first one is between $\phi_{k_1}^{128}(R_t^{cc=1\ldots m})$ and $\phi_{k_2}^{128}(R_{t-1}^{c=1\ldots n})$ to identify regions with state '*Normal HS*' in case of $\phi_{k_1}^{128}(R_t^{cc})$ matches to only one $\phi_{k_2}^{128}(R_{t-1}^{c})$ or prevent merging interaction ('*Merge*') in case of $\phi_{k_1}^{128}(R_t^{cc})$ matches to two or more $\phi_{k_2}^{128}(R_{t-1}^{c_1,c_2..c_q})$. The second one is between $\phi_{k_1}^{128}(R_t^{cc=1\ldots m})$ and

$Stopped(h).\phi$. $Stopped(h).\phi$ corresponds to region of stopped objects in previous frames, thus, if they mach, $R_t^{cc=1...m}$ are in state '*Re-Moving*', otherwise they are in state '*Entry*'. The third matching is between $\phi_{k_2}^{128}(R_{t-1}^{c=1...n})$ and $\phi_{k_3}^{128}(R_t^{c=1...m})$ to identify stopped objects, otherwise means disappearance of $R_{t-1}^{c=1...n}$ ('Exit'). $\phi_{k_3}^{128}(R_t^{c=1...m})$ correspond to SIFT descriptors of $\beta_t^{2...5}(R_{t-1}^{c=1...n})$ projection onto current frame.

3.2 Long- Term Processing (LTP)

We recall that the LTP rule feedbacks objects (O_i) in $TrackingObject\{O_{i=1...ObjectCount}$ and their corresponding regions in each frame to update tracked object attributes. Spatiotemporal attributes and descriptors of tracked object ($Z_t^{O_i} = (\beta^{1...5}(O_i), Cor, \phi(O_i))$) are updated according to region\object association. The association between objects and their corresponding regions is based essentially on $R_{j=\{t-1,t\}}^{i=\{c=1...n,cc=1...m\}}.Cor$. In fact, attributes of objects in states *'Entry'*, *'Split'* and *'Normal VE'* are updated according to Equation (5). Objects in state *'Stopped'* are controlled by $Stopped(O_{j=1...h}).\phi$ and objects in state 'Exit' ($TrackingObject\{O_i.Z_t^{O_i}(Cor)\} == c$ AND $R_{t-1}^{c}.Cor = *$) are killed.

$$\begin{cases} if\ (TrackingObject\{O_i.Z_t^{O_i}(Cor)\} == R_t^{cc}.Cor) \text{ OR } (R_t^{cc}.Cor == *) \text{ Then} \\ TrackingObject\{O_i.Z_t^{O_i}(Cor)\} = cc \\ TrackingObject\{O_i.Z_t^{O_i}(\beta^{1...5})\} = \beta^{1...5}(R_t^{cc}) \\ TrackingObject\{O_i.Z_t^{O_i}(\phi)\} = \phi(R_t^{cc}) \end{cases} \quad (5)$$

Attributes of objects in merging region are hardly obtained since several objects shared the same region (*cf.* Fig. 4 (A)). To deal with this problem, we use template matching (*cf.* Fig. 4 (B)) based sum of squared difference to find 2D spatial attributes of each object (*cf.* Fig. 4 (C)), then, we compute their SIFT descriptors. Sum of squared difference is implemented using FFT (Fast Fourier Transform) based correlation.

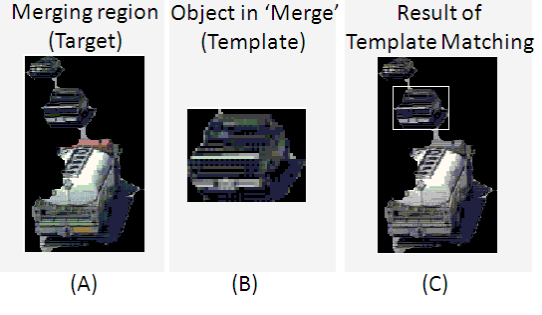

Fig. 4. Example of object localization in merging region. (A) Merging region (Target), (B) Object in 'Merge' and, (C) Result of Template Matching

4 Experimental Results

To evaluate our method, we used a corpus of 2 road traffic sequences[1] recorded in typical conditions (*HighwayII* and *HighwayIII*). *HighwayII* includes several interactions between road objects. *HighwayIII* include a dense traffic of road objects (different speed and size). The evaluation is made through the calculation of the rates of '*Centroid Error*' [13] regard to Ground-Truth (GT) of the two sequences parts (4 parts for each one). '*Centroid Error*' rates are computed according to two-pass matching scheme: first pass matching from system track to GT (*distanceSy*) to find false positive track (*FPT*) and second pass matching from GT to system track (*distanceTrack*) to find false negative track (*FNT*). In typical results, '*Centroid Error*' rates from the two pass are the same. In addition to the above quantitative metric, we also consider in our evaluation a second metric '*Two-pass many-to-many system to ground truth track matching*' [14] to measure how the system can deal with *'Merge'* and *'Split'* interactions. A GT/system track is matched to the system/GT track if there is both temporal overlap and spatial overlap. Temporal overlap is with respect to the duration of the system track. Spatial overlap is based on the centroid of the system lying inside the bounding box of the ground truth track. If multiple GT-matches, then this system track has '*Merge Error*' equal to matched GT tracks. If multiple system-matches, then this GT track has '*Split Error*' equal to matched system tracks.

As we can see in Fig.5, the four *HighwayII* parts (first line) echoed a low average *distanceSys/distanceTrack* rates peer part respectively between 0 and 6.163 pixels while *FPT* and *FNT* are between 0 and 6.19 percent. The four *HighwayIII* parts (second line) echoed a low average *distanceSys/distanceTrack* rates peer part respectively between 1.928 and 8.795 pixels while *FPT* and *FNT* are between 0 and 20,44%.

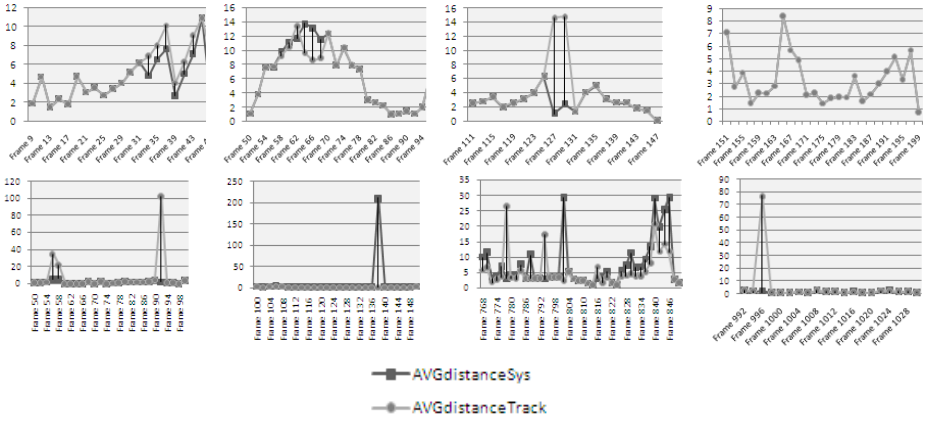

Fig. 5. Average distanceSys/distanceTrack curves of each frame of 4 part from *HighwayII* (first line) and *HighwayIII* (second line).

[1] Video sequences are courtesy of the Computer Vision and Robotics Research Laboratory at UCSD

We have performed the experimental study to know how our system can deal with *'Merge'* and *'Split'* interactions on 11 tracks from *HighwayII*. Temporal overlap and Spatial overlap curves for 4 of 11 tracks are depicted in Fig.6. For each track, both measures are computed firstly (A) from *GT-Track-Matching* and secondly (B) from *System-Track-Matching*. There is a '*Merge Error*'/'*Split Error*' in case of multiple GT-matches/system-matches, more explicitly, if a curve from *GT-Track-Matching/ System-Track-Matching* show more than peak with temporal overlap greater than 0.5 (*cf.* Track 4 for '*Merge Error*'). Our system achieves a '*Merge Error rate*' of 9.09 percent and a '*Split Error rate*' of 0 percent.

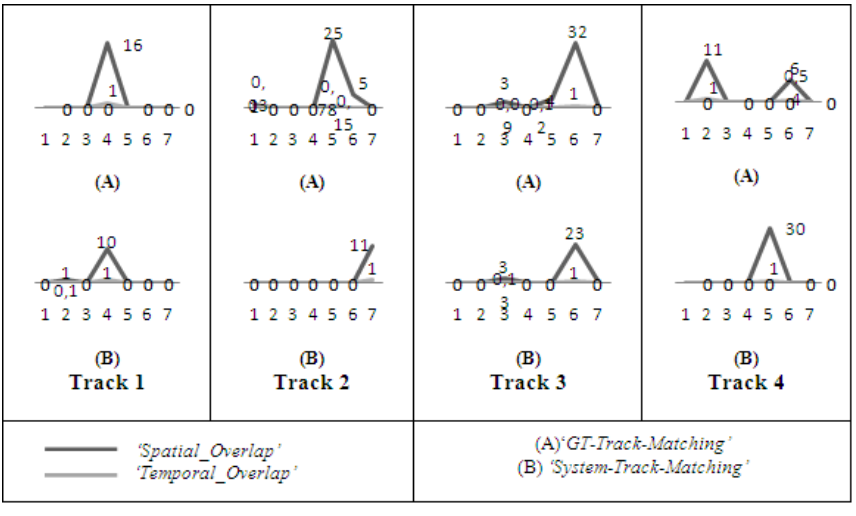

Fig. 6. Temporal overlap and Spatial overlap of 5 tracks from *HighwayII*

5 Conclusions

In this paper, we presented a novel method to track road objects. Our method is based on two prior processing: (1) Short-Term Processing (STP) that is based on spatial analysis and multilevel region descriptors matching**.** (2) Long- Term Processing (LTP) that is based on data association from STP. In these processing, both region and object information are used to establish objects correspondence over times.

The proposed algorithm was evaluated by a qualitative and quantitative experimental study on a corpus of road traffic sequences. The obtained results are rather satisfactory. In the near future, we plan to evaluate our method with computer vision applications like highway control and management system.

References

1. Peleshko, D., Ivanov, Y., Kustra, N., Kovalchuk, A.: An application of combined detector algorithm to extract the interest points of foreground objects in videostreams. In: 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics, p. 262 (2011)

2. Dan, L., Jian-sheng, Q.: Sift-based object matching and tracking of coal mine. In: IET 3rd International Conference on Wireless, Mobile and Multimedia Networks, pp. 327–330 (2010)
3. Lin, X., Zhang, J., Liu, Z., Shen, J.: Semi-automatic road tracking by template matching and distance transform. In: Joint Urban Remote Sensing Event, pp. 1–7 (2009)
4. Cremers, D., Schnörr, C.: Statistical shape knowledge in variational motion segmentation. Image and Vision Computing 21(1), 77–86 (2003)
5. Rahman, M., Saha, A., Khanum, S.: Multi-object Tracking in Video Sequences Based on Background Subtraction and SIFT Feature Matching. In: Fourth International Conference on Computer Sciences and Convergence Information Technology, pp. 457–462 (2009)
6. Yan, Y., Wang, J., Li, C.: Object tracking using SIFT features in a particle filter. In: IEEE 3rd International Conference on Communication Software and Networks (ICCSN), pp. 384–388 (2011)
7. Liu, Y., Wang, X., Yang, J., Yao, L.: Multi-objects tracking and online identification based on SIFT. In: International Conference on Multimedia Technology (ICMT), pp. 429–432 (2011)
8. Cheng-bo, Y., Jing, Z., Yu-xuan, L., Ting, Y.: Object tracking in the complex environment based on SIFT. In: 3rd International Conference on Communication Software and Networks, pp. 150–153 (2011)
9. Harris, C., Stephens, M.: A Combined Corner and Edge Detector. In: Alvey Vision Conference, vol. 15(Manchester), pp. 147–151 (1988)
10. Tomasi, C., Kanade, T.: Detection and Tracking of Point Features Technical Report CMU-CS-91-132, pp. 1–22 (1991)
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
12. Hammami, M., Jarraya, S., Ben-Abdallah, H.: On line Background Modeling For Moving Object Segmentation in Dynamic Scenes. Multimedia Tools and Applications Journal (available first on-line) (2011)
13. Senior, A., Hampapur, A., Tian, Y.-L., Brown, L., Pankanti, S., Bolle, R.: Appearance Models for Occlusion Handling. In: IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (2001)
14. Lisa, M.B., Andrew, W.S., Tian, Y.-L., Connell, J., Hampapur, A.: Performance Evaluation of Surveillance Systems Under Varying Conditions. In: IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (2005)