Proceedings of the 31st Chinese Control Conference
July 25-27, 2012, Hefei, China

Temporal Attention Control System for Multiple Objects Tracking

Xuelian Zang[1], Dražen Brščić[2], Wei Wang[3] and Kolja Kühnlenz[3,4]
1. Maximilians Universität München, 80802 Germany
E-mail: zangxuelian@gmail.com
2. Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan
3. Institute of Automatic Control Engineering,Technische Universität München, 80333 Germany
4. Institute for Advanced Study (IAS)

Abstract: An important ability for mobile robots is to process multiple tasks in complex environments. Since the sensor resources on a robot are limited, it is necessary to distribute the sensors attention to different tasks along the time scale. This paper proposes a temporal attention control method which aims at detecting multiple objects and estimating their poses with a single actuated camera. The proposed method is based on three criteria which are partially inspired by human behavior: (i) minimization of the overall object poses perception uncertainty and minimization of the variance of the perception uncertainty of different objects; (ii) minimization of the camera movements for completing the tasks; (iii) maximization of the number of objects in the cameras field of view.The proposed approach use Kalman filters to estimate object poses and to determine the perception uncertainty. The method was evaluated with both simulation and experiment on actual robot. The results show that the proposed approach is able to switch the camera's attention according to the objects poses and movements efficiently with a low frequency of the camera movements.

Key Words: Attention control, Kalman filter, multiple objects tracking

1 Introduction

Attention is the cognitive process of selectively concentrating on one aspect of the environment while ignoring the others. Similarly, computational visual attention systems aim to detect regions of interest. It is widely used in the research area of mobile robots [1–4]. With limited resources, a mobile robot system should be able to switch the sensors attention temporally for multiple tasks. The authors in [5] present a multi-camera view direction planning strategy to complete two concurrent tasks: robot self-localization and object tracking. In [6], an approach on tracking multiple moving objects with a mobile robot in populated environments is proposed and the tracking of moving objects with a mobile robot are extended in [7]. A goal-oriented attention guidance model is proposed in [8] to detect entities that are salient and relevant to the task. The method was introduced in [9] to track multiple moving targets using a camera and a laser range finder. However, the above mentioned works either used multiple sensors to achieve the multiple tasks or only dealt with the objects that exist in the cameras field of view (FOV). Approaches for selecting the viewpoint based on entropy were proposed in [10, 11]. A similar temporal attention control approach based on perception uncertainty minimization is introduced in [12]. However, the poses of the camera were limited to directions towards the objects.

Apart from the choice of the optimal direction, another important issue is the definition of the performance criterion. This paper proposes a temporal attention control system (TAC) with the following criteria: (i) the overall object poses estimation uncertainty minimization and the perception uncertainty variance minimization of different objects; (ii) minimizing camera movements for completing the task; (iii) maximizing the number of objects in the camera FOV. The possible camera direction is illustrates in Fig.1 (In this paper we consider a 2D tracking problem, but the results can be easily generalized to the 3D case). The paper is organized as follows: Section 2 introduces the related psychological models and the system model. Details of the proposed algorithm are described in Section 3. Experimental results are discussed in Section 4. Conclusions are given in Section 5.

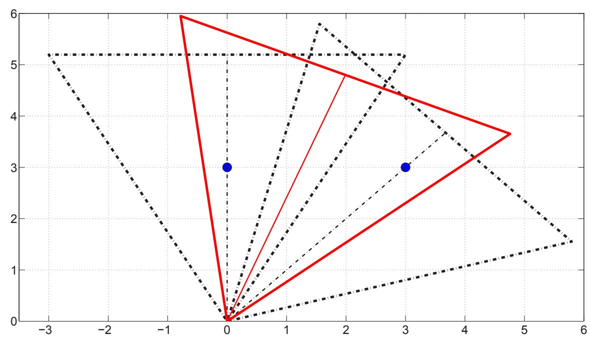

Fig. 1: Temporal attention control: due to the limited field of view, the camera direction is switched to achieve an optimal tracking performance. Some possible directions and fields of view: towards a single object (black dash dot line); covering multiple objects (red solid line).

2 Insight from human behavior and proposed approach

2.1 Temporal attention control in humans

Human achieve a high efficiency when processing multiobject tasks based on different strategies. The author in

This work is partly supported by the DFG excellence initiative research cluster Cognition for Tech- nical Systems CoTeSys, see www.cotesys.org, the FP7 EU-STREP Interactive Urban Robot (IURO), see www.iuro-project.eu, the Institute for Advanced Study (IAS), Munich and the Chinese Scholarship Council.

[12] present an experiment on human eye movements when considering various relevant aspects for attention planning. In the experiments, three participants first stood randomly around an object. They were asked to move around the object and distribute themselves equally. The participants' movements and gaze switching were recorded and analysed. The results showed that humans try to minimize the perception uncertainty for observing objects according to multi-object tasks.

Human is able to track multiple objects at the same time was presented in [13–15] . For the experiments in [14], subjects were asked to track multiple items as the objects moved independently and unpredictably about the display. The results showed that the performance to track multiple objects was not impaired even when the items were briefly (but completely) occluded at various times during their motion. This suggested that human are not only able to track multiple objects at the same time, but also able to track and predict the locations when they were briefly occluded. It was proposed that a mechanism with both parallel and serial processing and temporary spatial memorywas involved in the multiple object tracking in [15] according to their experiments. They examined the tracking task with both identical and distinct objects (visually and semantically different), and the results showed that tracking performance deteriorated as a function of tracking time and set size.

Psychology experiments on switching cost for multitasking are described in [16–18]. Switching-time cost for multiple tasks such as solving math problems or classifying geometric objects were presented in [16]. It is found that the switching time cost can be reduced according to the complexity of the tasks but not eliminated. Effects of task and location switching on the accuracy of the reporting target characters in an attentional blink paradigm are examined in [17]. The perceived accuracy is affected by both task switching cost and location switching. According to these results, the optimal attention control should consider the minimum of switching in order to have less switching time cost and increase the perception accuracy.

2.2 Proposed method

Inspired by the human behavior results presented above, we propose a single camera attention control method which control the robot worked in a more efficiency way. Three criteria were considered as follows: (i) minimization of overall perception uncertainty and minimizing the variance of perception uncertainty of different objects; (ii) maximization of the number of objects in the camera field of view; (iii) minimization of camera movements.

The structure of the system proposed in this paper for achieving the above requirements is shown in Fig. 2. It includes the initialization, object detection and poses estimation and TAC. In the initialization step, the camera is controlled to scan the world to search and compute the initial poses of the task related objects. Object detection and pose estimation are prerequisites of the system which provides the measurement of the objects at different time steps. In this paper, however, we will focus on the TAC which is the main contribution of our work. TAC includes two steps: (i) Optimal camera direction computation step with the aim of the current object poses estimation and computation of the optimal camera angle for the next time step; (ii) Switching attention step with the purpose of changing the cameras direction to the desired value for capturing new information from the environment. To switch between these two steps, we proposed a strategy to control the camera states. Kalman filter is used to estimate the object poses from the measurement data and compute the perception uncertainty. Details are discussed in the next section.

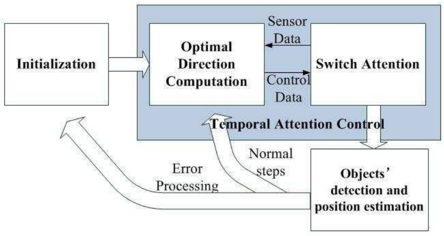

Fig. 2: Structure of the proposed system

3 Temporal attention control strategy

3.1 Camera states control strategy

The camera states control strategy is designed as the switching between optimal camera direction computation step and switching attention step. Based on a performance criterion, only images captured in the first step are processed. We defined the switching criterion according to the overall perception uncertainty of objects: If the overall perception uncertainty of the objects in the camera's FOV is low enough, or the overall perception uncertainty of objects out of camera's FOV is large enough, the system will switch to the second step. Otherwise, the system will stay at the optimal camera direction computation step to improve the estimation results for the objects in the camera's current FOV. The switching condition for the TAC strategy is:

$$Sign = \begin{cases} 1 & \text{if } \left|p_{seen_k} - p_{seen_{k-1}}\right| \leq Th_1 \\ & \text{ or } p_{unseen_k} \geq Th_2 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The system keeps at switching attention step when $Sign = 1$, and keeps at the optimal camera direction computation step when $Sign = 0$; p_{seen_k} is the perception uncertainty of the objects in the current cameras FOV, and p_{unseen_k} is the perception uncertainty of the objects that are out of the current camera's FOV. Th_1 and Th_2 are thresholds for perception uncertainty.

3.2 Kalman filter for perception uncertainty estimation

We use a Kalman Filter (KF) which is parameterized by measurement noise covariance R and the process noise covariance Q to estimate perception uncertainty [12, 19]. The system states are chosen as poses and velocities of the objects $X = (x\ y\ z\ \dot{x}\ \dot{y}\ \dot{z})^T$.

We assume that object j's measurement noise, i.e. R_j depends on the camera properties and the object poses in the camera coordinates [12]. It decreases linearly with the distance from the object to the optical axis and the distance between the object and the camera.

$$R_j = \begin{pmatrix} k_{c1}x_{jc} & k_{c2}y_{jc} & k_{c3}z_{jc} \end{pmatrix}^T \tag{2}$$

where k_{c1}, k_{c2}, and k_{c3} are the parameters that express the relationship between the measurement error covariance and the objects poses and $\begin{pmatrix} x_{jc} & y_{jc} & z_{jc} \end{pmatrix}^T$ is the object j's position in the camera's coordinate.

The process noise of object j, i.e. Q_j is biased by the speed and acceleration of the objects' movements. It is defined as in [12]:

$$Q_j = \begin{pmatrix} Q_{j,k,pos} & 0_n \\ 0_n & Q_{j,k,vel} \end{pmatrix} \tag{3}$$

where $Q_{j,k,pos}$, $Q_{j,k,vel}$ are defined as,

$$Q_{j,k,pos} \propto \dot{X}_{jc} \; and \; Q_{j,k,vel} \propto \ddot{X}_{jc}. \tag{4}$$

where $X_{jc} = \begin{pmatrix} x_{jc} & y_{jc} & z_{jc} \end{pmatrix}^T$.

The perception uncertainty of object j is:

$$E_j = \sum_{l=1}^{n} \sqrt{e_{j,l}^2} \tag{5}$$

where $e_{j,l}$ are the first n eigenvalues of the covariance matrix P_j, $n = 3$ is the dimension of the objects position and $l = 1...n$ denotes x, y, z directions.

3.3 Computing the optimal camera direction

Computing the optimal camera direction for the next time step is one of the key issues of TAC. To simplify the computation, instead of searching in the continuous space of all possible directions, we first compute a discrete set of direction candidates. Then the final optimal direction is computed among these candidates.

3.3.1 Computing the direction candidates

The possible directions are chosen as follows: First of all, the directions where the camera faces directly towards the objects are important since the measurement error of the current attended object is low. The second set of the possible directions are the directions that cover several objects at the same time if these objects are close enough. The last candidate is the current camera direction which aims at minimizing the camera movements. The whole set of candidates $\Omega_{k+1|k}$ can be described as:

$$\Omega_{k+1|k} = \begin{cases} \Omega_k^0 \\ \Omega_{argmin}(\sum_{j=1}^{m} d_{jc}) \\ \Omega_k^* \end{cases} \tag{6}$$

where Ω_k^0 includes the angles that the camera directly faces the objects, $\Omega_{argmin}(\sum_{j=1}^{m} d_{jc})$ considers the angles that cover several objects in FOV at the same time, and Ω_k^* includes the optimal angle from the last time step. $j = 1...m$ is the number of the objects that exist in the cameras FOV, d_{jc} is the distance from the object j to the cameras optical axis.

3.3.2 Computing the final optimal camera direction

This part aims at finding the final optimal direction $\Omega_{k+1|k}^*$ for the time step $k + 1$ from the optimal direction candidates. According to the above description, $\Omega_{k+1|k}^*$ is computed from:

$$\Omega_{k+1|k}^* = \underset{\Omega_{k+1|k}}{\operatorname{argmin}}(J) \tag{7}$$

where the objective function J is composed by four sub-objective functions J_1, J_2, J_3 and J_4 which are described in the following part. Before weighting the sub-objective functions, they are normalized.

$$\begin{aligned} J(\Omega_{k+1|k}) = w_1 \cdot Norm(J_1(\Omega_{k+1|k})) \\ +w_2 \cdot Norm(J_2(\Omega_{k+1|k})) \\ +w_3 \cdot Norm(J_3(\Omega_{k+1|k})) \\ +w_4 \cdot Norm(J_4(\Omega_{k+1|k})) \end{aligned} \tag{8}$$

The overall perception uncertainty J_1 is defined as [12, 20]:

$$J_1(\Omega_{k+1|k}) = \frac{1}{n}\sum_{j=1}^{m} E_j(\Omega_{(k+1|k)}) \tag{9}$$

where $E_j(\Omega_{(k+1|k)})$ is the perception uncertainty of object j and m is the number of task-relevant objects. Only considering the minimization of the overall perception uncertainty, the system could reach the state where a few object perception uncertainties are very high while the others are very low. In this way, the overall perception is still low enough to be taken as the optimal view direction for the next time step. However the system may lose the information of the objects with high perception uncertainty. It is very important to manage all the object poses estimation in a good stage at the same time when facing the multiple objects attention planning problem. Considering this issue, function J_2 aims at minimizing the variance of the perception uncertainty of the objects:

$$J_2(\Omega_{k+1|k}) = VAR(E(\Omega_{k+1|k})) \tag{10}$$

where

$$E(\Omega_{k+1|k}) = \left(E_1(\Omega_{k+1|k}), E_2(\Omega_{k+1|k}) \cdots E_m(\Omega_{k+1|k})\right)^T \tag{11}$$

As described in the previous section, the TAC should also include the minimization of attention switching and switching cost. J_3 aims at minimizing the cameras movements to realize the switching minimization:

$$J_3(\Omega_{k+1|k}) = \left|\Omega_{k+1|k}^* - \Omega_{k+1|k}\right| \tag{12}$$

The view coverage function J_4 is defined to maximize the number of the objects in the cameras FOV. It is the ratio of the number of objects located in the cameras view (M_{seen}) and the number of the task related objects (M):

$$J_4(\Omega_{k+1|k}) = -M_{seen}/M \tag{13}$$

The weighting parameters w_1, w_2, w_3, w_4 in equation(8) ($w_1 + w_2 + w_3 + w_4 = 1$) are used to define the influence degree of each criterion on the overall performance. In our experiments, we compute an adaptive setting of the weights according to average values of overall perception uncertainty $J_1(\Omega_{k+1|k})$ and the variance of perception uncertainty $J_2(\Omega_{k+1|k})$ (expressed as $\bar{J}_1$ and $\bar{J}_2$):

(i) When $\bar{J}_1$ is too large, increase w_1 and decrease the other parameters, in this way, the object function J is mainly biased by J_1 which aims at minimize the perception uncertainty; (ii) When $\bar{J}_1$ is small but $\bar{J}_2$ is large, w_2 should be increased to balance the perception uncertainty of different objects; (iii) When both $\bar{J}_1$ and $\bar{J}_2$ are small enough, the perception results is acceptable. w_3 and w_4 should be increased to minimize the camera movements and enlarge the number of the objects in the camera's FOV.

The final optimal camera direction is chosen from the optimal angle candidates according to the above functions.

4 Simulation and experiments

The proposed approach was tested both in simulations in Matlab and in an implementation on a system with a camera mounted on an actuation unit.

As described in Section 3, the performance of the proposed approach varied by the value of weights (w_1, w_2, w_3, w_4). An adaptive way of choosing weights for simulations is presented (shown in Table 1). In the table, the columns show different condition of $\bar{J}_1$ and the rows show different condition of $\bar{J}_2$. The thresholds according to the simulation environments are defined as $th_1 - th_7$. The other parts in the table show the value of w_1 and w_2 according to $\bar{J}_1$ and $\bar{J}_2$. w_3 and w_4 are computed by $w_3 = w_4 = (1 - w_1 - w_2)/2$.

Table 1: Weights Computation

(w_1, w_2)	$\bar{J}_2 \geq th_2$	$th_2 \geq \bar{J}_2 \geq th_1$	$th_1 \geq \bar{J}_2 \geq 0$
$\bar{J}_1 \geq th_7$	(0.6 0.4)	(0.6 0.3)	(0.6 0.2)
$th_7 \geq \bar{J}_1 \geq th_6$	(0.5 0.4)	(0.5 0.3)	(0.5 0.2)
$th_6 \geq \bar{J}_1 \geq th_5$	(0.4 0.4)	(0.4 0.3)	(0.4 0.2)
$th_5 \geq \bar{J}_1 \geq th_4$	(0.3 0.4)	(0.3 0.3)	(0.3 0.2)
$th_4 \geq \bar{J}_1 \geq th_3$	(0.2 0.4)	(0.2 0.3)	(0.2 0.2)

4.1 Simulation with static objects

A similar attention planning (AP) is proposed in [12] which also aims at minimizing the overall perception uncertainty. However, AP did not consider the minimization of the variance of the perception uncertainty and could not attend to several objects at the same time. To compare with attention planning (AP), a simulation similar to the one presented in [12] is designed to detect four static objects in the environment.

Objects O_1, O_2, O_3 and O_4 are located at $(1.5, 4)$m, $(1.75, 3)$m, $(3, 2)$m, $(0.5, 1.74)$m, and the camera is located at the zero point $(0, 0)$m. Random Gaussian noise is added on the objects poses to simulate the actual environment. The weights parameters are set according to the adaptive approach described above.

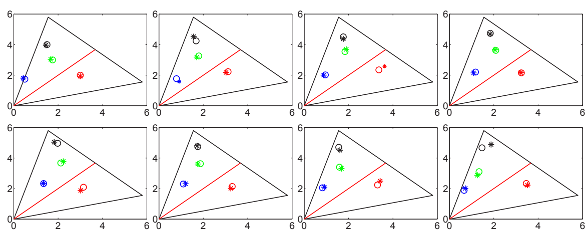

Fig. 3: Simulation results with static objects. Circles: the actual objects poses; Stars: the predicted objects pose

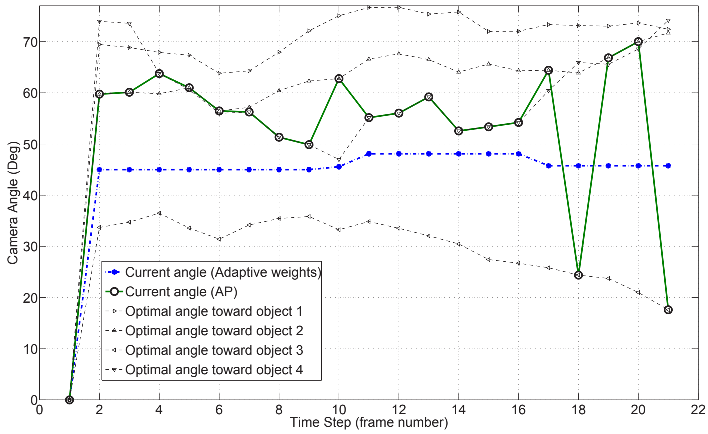

Fig. 4: Camera angle of simulation with static objects: Blue star line: actual pan angle with TAC; Green circle line: actual pan angle with AP; Other lines: the directions when the camera faces directly towards objects.

The first eight time step of the cameras observations (the sequence of the images is from up-left to the down-right and the triangles represent the cameras FOV) was shown in Fig.3. The camera direction barely changes in this simulation. Fig. 4 shows the actual attention direction (expressed as pan angles) of 20 time steps and compared with the results from AP. The blue star line is the current attention direction with TAC. The green circle line is the current attention angle with the AP method, and the other four lines are the directions where the camera directly points towards the objects. Table 2 shows the mean and variance of the pose estimation error and Table3 shows the number of camera switching times and the overall camera movements distance in the whole time scale. The results show that the attention switches less and moves smaller with TAC than with AP. The poses estimation uncertainty is also smaller.

Table 2: Poses estimation error of simulation with static objects

	x direction		y direction	
	mean[m]	var[m]	mean[m]	var[m]
TAC	0.48	0.05	0.42	0.03
AP	0.47	0.09	0.55	0.24

Table 3: Camera movements of simulation with static objects(Switch Times: The number of times for the camera switching; MoveDist: The overall move distance of the camera in the whole time scale)

	Switching Time	Move Dist(Degree)
TAC	3	50.90
AP	20	173.41

4.2 Simulation with dynamic objects

In order to evaluate the dynamic performance of the approach, we also tested the system with dynamic objects. Fig.5 shows the objects movements in the dynamic simulation scenario. This simulation takes 20 time steps. The arrows in the image show the movement directions of each object while the data in the brackets shows the speeds in the current situation. For example, object 1 (black solid line) first moves at the speed 1.0 m/frame in the x direction and 0.5m/frame in the y direction, after a certain period of time (5 frames), the speed changes to -0.5m/frame in the x direction, and 0.25m/frame in the y direction, then the speed changes again. Noises are added to the object movements.

To compare the results of the proposed approach, different simulations were done with AP method and TAC with varied weights. Fig. 6 shows the camera angles at different time steps with different methods. The black dashed lines with triangles shows the angles for the camera faces directly towards the objects. The green solid line with circles shows the current camera directions computed by AP method. The blue dash dot bold line shows the current camera direction using TAC method with adaptive weights while the red dashed line and the black solid line shows the directions with fixed weights ($(w1; w2; w3; w4) = (1; 0; 0; 0)$ and $(w1; w2; w3; w4) = (0.5; 0.5; 0; 0)$). Table 4 shows the mean and variance of the overall estimation error of the four objects in X and Y directions over 20 time steps (AW: simulation using TAC method with adaptive weights, FW1: simulation using TAC method with fixed weights as $(w1; w2; w3; w4) = (1; 0; 0; 0)$; FW2: TAC with fixed weights as $(w1; w2; w3; w4) = (0.5; 0.5; 0; 0)$). Table 5 shows the sum estimation error for both X and Y directions. Table 6 shows the number of times of the camera attention switching and the overall moving distance in the whole time scale (Switch Times: The number of the camera switching times; Distance(Deg): The overall move distance of the camera in the whole time scale). TAC shows a better performance than AP in the following aspects: (i) TACneed less number of times for the camera movements and less overall movements distance. (ii) The means and variances estimation error are lower with TAC.

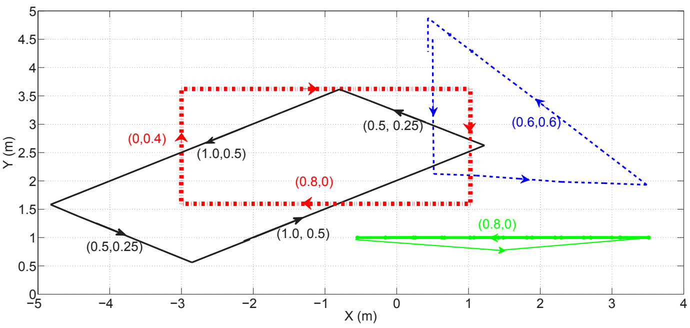

Fig. 5: Object's movements

As described in Section 3, TAC performs differently with different weights. With adaptive weights, TAC performed optimally with both lower estimation error and lower camera movements. When only considering J_1 (the overall perception uncertainty), the tracking results of the system is better than AP, but worse than when both minimization of J_1 and J_2 are considered.

4.3 Experiment on actual robot

We tested the method on an experimental setup using a PointGray Bumblebee XB3 camera (with 66^o FOV) which was mounted on a pan-tilt unit. A marker-based tracking task was used to evaluate the system shown as Fig. 7. Markers for the experiment were placed in such a way that the camera could not capture all the markers at the same time. The detection of the markers was implemented using the [21]. The sub-images in the right corner of Fig. 7 show the images captured by the camera at different time steps. From these two images, we can see that the system needs to switch the cameras directions from time to time to monitor all the three markers in the environment. The results show that the proposed system provides a good performance with a low perception uncertainty as well as a low frequency of attention switching.

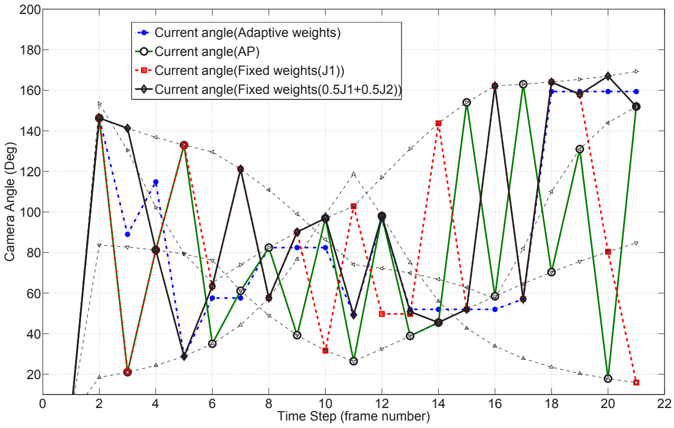

Fig. 6: Camera angles for simulations with varied weights: Blue dash dot bold line with stars: actual camera angle with adaptive weights; Red dashed line with squares: camera pan angle with fixed weights ($f = J_1$); Black solid line with diamonds: actual camera angle with fixed weights ($f = 0.5J_1 + 0.5J_2$); Black lines with triangles: the directions when the camera faces directly towards objects.

Table 4: Poses estimation error for dynamic simulations with varied weights (AW: TAC with adaptive weights; FW1: TAC with fixed weights as $(w_1, w_2, w_3, w_4) = (1, 0, 0, 0)$; FW2: TAC with fixed weights as $(w_1, w_2, w_3, w_4) = (0.5, 0.5, 0, 0)$)

Method		x direction		y direction	
		mean[m]	var[m]	mean[m]	var[m]
AP		2.17	2.03	1.77	1.23
TAC	AW	1.11	0.67	1.21	1.28
	FW1	1.78	1.19	1.50	1.34
	FW2	1.44	1.11	1.67	1.30

4.4 Conclusion

A temporal attention control strategy to monitor multi-objects is proposed. The overall perception uncertainty of the objects is estimated by Kalman filter. To provide an optimal control, four sub functions are considered to optimize the perception objective and camera movements. An adaptive way for computing the weights is given in the simulation part. Simulations and experiment on robots were performed to test the proposed method. The results show a good performance with a low perception uncertainty and low energy cost for switching attention.

Table 5: The overall mean of poses estimation error for dynamic simulations with varied weights

	AP	TAC(AW)	TAC(FW1)	TAC(FW2)
Error	3.94	2.32	3.27	3.11

Table 6: Camera movements for dynamic simulation with with varied weights (Switch Times: The number of times for the camera switching; MoveDist: The overall move distance of the camera in the whole time scale)

Method		Switch Times	Move Dist(Degree)
AP		20	1471.3
TAC	AW	11	598.7
	FW1	19	1375.1
	FW2	20	960.7

References

[1] Simone Frintrop and Markus Kessel, Most Salient Region Tracking, *IEEE international conference on Robotics and Automation (ICRA)*, 2009.

[2] Daniel Stronger and Peter Stone, Selective Visual Attention for Object Detection on a Legged Robot, *RoboCup*, 2006

[3] Tingting Xu, Tianguang Zhang, Kolja Kühnlenz and Martin Buss,Attentional Object Detection of An Active Multifocal Vision System, *International Journal of Humanoid Robotics(IJHR)*,Volume 7, Issue 2. 2010

[4] Simone Frintrop,Computational visual attention systems and their cognitive foundations: A survey, *ACM Transactions on Applied Perception*,Volume 7, Issue 1. 2010

[5] Tingting Xu, Koljia Kühnlenz and Martin Buss, A View Direction Planning Strategy for a Multi-Camera Vision System, *Proceedings of IEEE International Conference on Information and Automation (ICIA)*, 2008

[6] Dirk Schulz, Wolfram Burgard, Dieter Fox and Armin B. Cremers, Tracking Multiple Moving Objects with a Mobile Robot, *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*,2001

[7] Andreas Kräußling,Tracking Extended Moving Objects with a Mobile Robot, *Intelligent Techniques and Tools for Novel System Architectures, studies in Computational Intelligence*, 2008

[8] Vidhya Navalpakkam and Laurent Itti, A Goal Oriented Attention Guidance Model, *Lecture Notes in Computer Science*, 2002

[9] Boyoon Jung and Gaurav S. Sukhatme, Tracking Multiple Moving Targets using a Camera and Laser Rangefinder, *Institutue for Robotics and Intelligent Systems (IRIS) Technical Report IRIS-01-397 University of Southern California*, 2001,

Fig. 7: Experiment on actual camera

[10] Pere-Pau Vázquez, Miquel Feixas, Mateu Sbert and Wolfgan Heidrich, "Viewpoint Selection using Viewpoint Entropy", *Proceedings of the Vision Modeling and Visualization Conference*, 2001

[11] Tal Arbel and Frank P. Ferrie,Viewpoint Selection by Navigation through Entropy Maps, *Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV)*, 1999

[12] Tingting Xu,Aspects of Visual Attention for Autonomous Mobile Robots, PhD thesis, *Technische Universitaet Muenchen, Munich, Germany*, 2010

[13] Christopher R. Sears and Zenon W. Pylyshyn, Multiple Object Tracking and Attentional Processing, *C*anadian Journal of Experimental Psychology, 2000

[14] Brain J. Scholl and Zenon W. Pylyshyn,Tracking Multiple Items Through Occlusion: Clues to Visual Objecthood, *Cognitive Psychology*, Volume 38, Issue 2, 1999

[15] Lauri Oksama and Jukka Hyönä,Is multiple object tracking carried out automatically by an early vision mechanism independent of higher order cognition? An individual difference approach, *Vision Cognition*, Volume 11, Issue 5, 2004

[16] Joshua S. Rubinstein, David E. Meyer and Jeffrey E. Evans, Executive Control of Cognitive Process in Task Switching, *Journal of Experimental Psychology: Human Perception and Performance*, Volume 27, Issue 4, 2001

[17] Stephen Monsell,Task switching, *Trends in Cognitive Sciences*, 2003

[18] James F. Juola, Juan Botella and Antonio Palacios,Task-and location-switching effects on visual attention, *Perception & Psychophysics*, 2004

[19] Welch Greg and Bishop Gary,An Introduction to the Kalman Filter, http://www.cs.unc.edu/ welch/kalman/index.html, 1995

[20] J.F.Seara,Intelligent Gaze Control for Vision-Guided Humanoid Walking, PhD thesis, *Technische Universitaet Muenchen, Munich, Germany*, 2004

[21] ARTOOLKIT, http://www.washington.edu/artooklit/, 2002