15

Active Tracking System with Rapid Eye Movement Involving Simultaneous Top-down and Bottom-up Attention Control

Masakazu Matsugu, Kan Torii and Yoshinori Ito
Canon Inc.
Japan

1. Introduction

Visual tracking of complex objects has been studied extensively in the field of robot vision, visual servoing, and surveillance. Detection of reliable low level visual features has been crucial, in the literature, for the stability of tracking in cluttered scene. However, using only low level features for tracking tends to be illusory for practical vision systems, and the selection of reliable visual features is still an important unsolved issue.

In recent studies it has been stressed the importance of high-level features in guiding control for long-duration tracking (Matsugu et al., 2006; Li et al., 2007; Yang et al., 2007), computer vision (Sun & Fisher, 2003), and visual search (Lee et al., 2005). Recent cognitive neuro-psychological as well as brain imaging studies also revealed a role of top-down attention in visual search task in human vision system (Patel & Sathian, 2000; Hopfinger et al., 2000; Corbetta & Shulman, 2002; Navalpakkam & Itti, 2006).

In this chapter, we present a new object tracking vision system that incorporates both top-down and bottom-up attention processes. Tracking a specific object using both low-level and high-level features is not new and was previously studied (Isard & Blake, 1998) in a stochastic framework. The proposed active vision system in this chapter is task-oriented and tracks a specific object (i.e., person). The tracking process is initiated by the top-down process which is activated by robust object detection module (Matsugu & Cardon, 2004; Matsugu et al., 2004). Subsequent feature selection for tracking involves simultaneous *consolidation* mechanism between higher level complex features (e.g., face) obtained from the top-down process and low level features from the bottom-up process (detailed description is in Section 3). The active vision system controls eye movement based on prediction of an attended object's location by the above processes.

Main contribution of this chapter is that we propose a stable object tracking algorithm involving selective attention together with FF and FB hybrid control that enable smooth and saccadic pursuit. Specifically, we introduce a coherency measure of tracking features. Tracking using such measure ensures stability and fast recovery from failure (missing the object to be tracked) by way of *consolidation* among bottom-up and top-down attention cues.

For the bottom-up feature-based prediction of tracked object, we use local color histogram, and histogram intersection (Swain & Ballard, 1991; Birchfield & Rangarajan, 2005) is used for feature matching. High-level, top-down feature for tracking is defined as detected face

Source: Computer Vision, Book edited by: Xiong Zhihui,
ISBN 978-953-7619-21-3, pp. 538, November 2008, I-Tech, Vienna, Austria

location and its predicted location. The coherency measure is used also for feature selection resulting in switching in-between low-level and high-level tracking features, and the active vision system controls panning/tilting angle of camera module.
We demonstrate fast and stable tracking of a person moving rapidly with abrupt motion change, attaining maximum instantaneous panning speed of 6deg/40ms. This saccadic pursuit is roughly equivalent to tracking an object moving at 10m/sec at the distance of 3m. Persistent tracking with stability and robustness is demonstrated in a number of experiments under drastic illumination changes, background clutters, and occlusions.
The organization of this chapter is as follows. In Section 2, we review conventional active vision systems for object tracking. In Section 3, details about the proposed algorithms are given and we propose a new feature coherency measure for selecting useful features for tracking, followed by demonstration of robustness and stability of tracking as well as fast and smooth eye movement control in a number of human tracking experiments. In Section 4, details about hardware implementation are explained. Section 5 shows results and performance, followed by discussion and conclusions in Section 6 and 7.

2. Overview of attention control, object tracking, and active vision

A great many models for attention control have been proposed (e.g., Koch & Ullman, 1985; Itti & Koch, 2000). Most existing algorithms in machine vision, with a few exceptions (e.g., Culhane & Tsotsos, 1992; Yang et al., 2007), exploit a class of saliency map derived from low-level features (Koch & Ullman, 1985; Treisman & Gelade, 1980) and a spatial attention window operates on the saliency map. It is a map of scalar values that encodes locations of something conspicuous in the visual field to guide vision related tasks such as visual search and servoing. Most of saliency measure is described in terms of low-level features such as color cues (Swain et al. 1992) or motion cues (Cretual et al., 1998; Bur et al., 2007) that are *different and popped-out* from neighborhood surrounding features.
On the other hand, a limited number of methods have been proposed for attention control that involves both top-down and bottom-up processes (Giefing et al., 1992; Olshausen et al., 1995; Sun & Fisher, 2003; Lee et al., 2005; Yang et al., 2007). Interest map (Giefing et al., 1992) is a kind of saliency map which unifies maps of pre-attentive low-level features and attentive features indicating object hypothesis, which is used to induce top-down saccadic shift. Dynamic routing circuit (Olshausen et al., 1995) was introduced for top-down attention control in a hierarchical feed-forward neural network model for object detection. Late attentional selection mechanism (Yang et al., 2007) involves top-down, higher cognitive process to enhance robustness of tracking under occlusion and cluttering and also to ensure long duration tracking. In this model, low-level features are used for bottom-up fast process and high-level features for top-down, slow process. A slow attentional process using top-down, object cue (Sun & Fisher, 2003) was combined with a fast, bottom-up process. Biologically motivated attention control architecture (Mancas et al., 2007) is also proposed with three levels of units, namely, low-, mid-, and high-levels, as opposed to conventional two levels of pre-attentive and attentive systems.
Integration of multiple modalities has been proven to be helpful for enhancing robustness in tracking as they act complementary source of visual cues. In fact, a saliency map is a class of integral representation of multiple cues (Itti et al., 1998; Bur et al., 2007). In particular, the integration of different cues for tracking using stochastic framework has long been addressed in the literature (Isard & Blake, 1998; Rasmussen & Hager, 2001; Wu & Huang,

2001, Serby et al., 2004). For instance, multiple low-level features (e.g., edges, textures, interest points) can be integrated into a particle filter framework for tracking (Serby et al., 2004). Covariance tracking (Porikli et al., 2006) can also combine different modalities such as infrared and color cues in covariance matrix. Integration of detection and tracking has also been proven to be effective for enhanced reliability (Sigal et al., 2004; Matsugu et al., 2006; Li et al., 2007). Model-based cues and model-free cues are also integrated in the framework of Kalman filtering (Kyrki & Kragic, 2005).

Complex object tracking (Drummond & Cipolla, 2002) is another important issue in robot vision functionalities such as visual servoing (Clark & Ferrier, 1992; Cretual, et al. 1998; Fujita et al., 2007), random bin-picking in eye-in-hand system, and people tracking in surveillance (Comaniciu et al., 2000; Li et al., 2006; Yang et al., 2006; Matsugu et al., 2006). In the field of visual search strategy for traditional purposive vision (Garvey, 1976; Ballard & Brown, 1992), fixation or gaze control was realized as a manifestation of attention control. The task oriented visual search deals with selecting objects and order of fixation that result in the control of eye movement.

Active vision system with tracking functionality has been studied extensively in the literature. A number of methods on object tracking with camera motion have been proposed (Hunt & Sanderson, 1982; Aloimonos et al., 1987; Burt et al, 1989; Birchfield, 1997; Birchfield & Rangarajan, 2005; Blake & Yuille, 1992; Murray & Basu, 1994; Bradshaw et al, 1994; Castrillón-Santana et al., 1998; Comaniciu et al., 2000; Murao et al., 2006; Yang et al., 2006), however, only a few works addressed co-realization of saccadic, quick eye-movement and smooth pursuit in an active vision head with pan/tilt operation. A notable exemplary work is coarse and fine resolutions for saccadic and smooth pursuit (Bradshaw et al., 1994).

As regards feature selection in active tracking system that involves pan/tilt/zoom control, primitive features such as color and motion have been favorably used in the literature. A notable system for tracking specific object using color cue is mean shift method (Comaniciu et al., 2000) that relies on color histogram for camera control. On-line feature selection is also important for enhanced stability. In a face tracking active vision (Yang et al., 2006) dominant color feature was selected during tracking. Motion cue has also been used as major and important feature. For instance, majority tracking algorithm (Burt et al., 1989) used optical flow and was implemented on pyramid vision hardware. Motion energy and ego-motion compensation (Murray & Basu, 1994) was introduced for calculating pan/tilt angles about the lens center. As in attention control, there have been a limited number of works (Castrillón-Santana et al., 1998, Matsugu et al., 2006; Yang et al., 2007) on using both low level and high level cues in active vision systems.

3. Tracking with selective attention

3.1 Top-down process

Specific goal of our task oriented active vision system is *long duration tracking* of a person under drastic illumination change, in cluttered scene with lots of distracters. The top-down process which constitutes slow subsystem involves fast and robust object detection (Matsugu & Cardon, 2004) using modified convolutional neural networks (MCoNN). In Fig. 2, we show a schematic architecture of MCoNN which constitutes a subsystem for face detection. The top-down stream generates prediction data sequence of a specific object to be tracked.

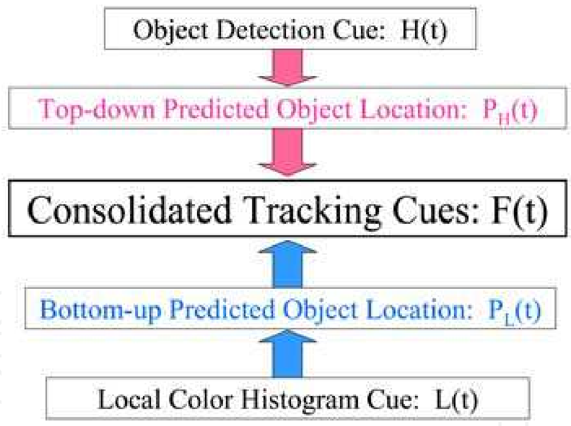


Fig. 1. Top-down and bottom-up attention cues are integrated as consolidated tracking cues

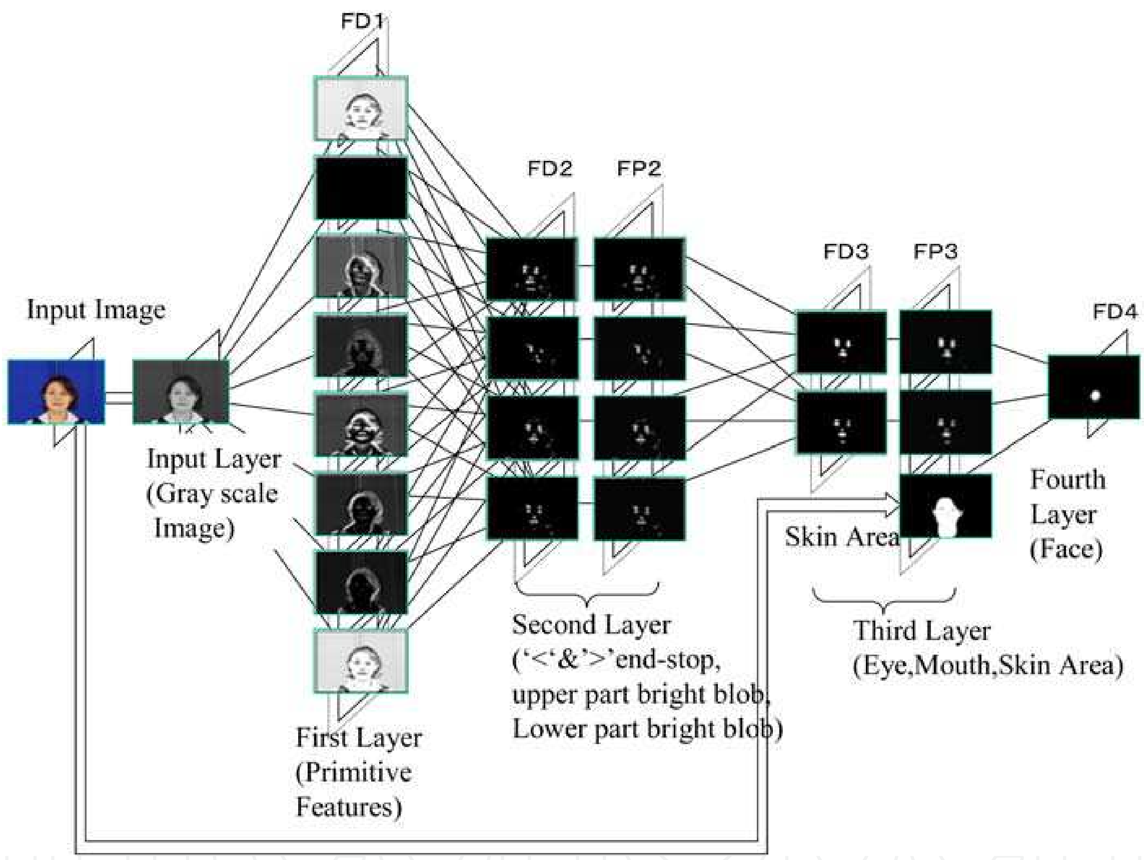


Fig. 2. Hierarchical architecture (MCoNN) for integrating predefined alphabetical local features (indicated in the second layer level) for object detection.

Fig. 3 shows a schematic diagram of object (face) detection algorithm as a whole. The algorithm is composed of two sub-modules. The first stage is a combination of modified cascaded filtering (Viola & Jones, 2001) system for fast and robust face detection, followed by MCoNN. The modified V&J (details will be published elsewhere) uses cascaded simple filters (e.g., rectangle filters extracting Haar-like features) and *integral image* representation (Crow, 1984; Viola & Jones, 2001). The second stage is composed of a set of MCoNNs each tuned to a specific rotation/size class of faces (Mitarai et al., 2003).
As indicated in Fig.1, sequence of locations of a particular object-to-be-tracked obtained from the top-down subsystem is stored and fed to an intermediate subsystem that evaluates

coherency. The intermediate subsystem receives inputs from bottom-up stream as well, and it integrates those cues from the two streams to predict the location of the object in the upcoming frames as *consolidated* tracking cues.

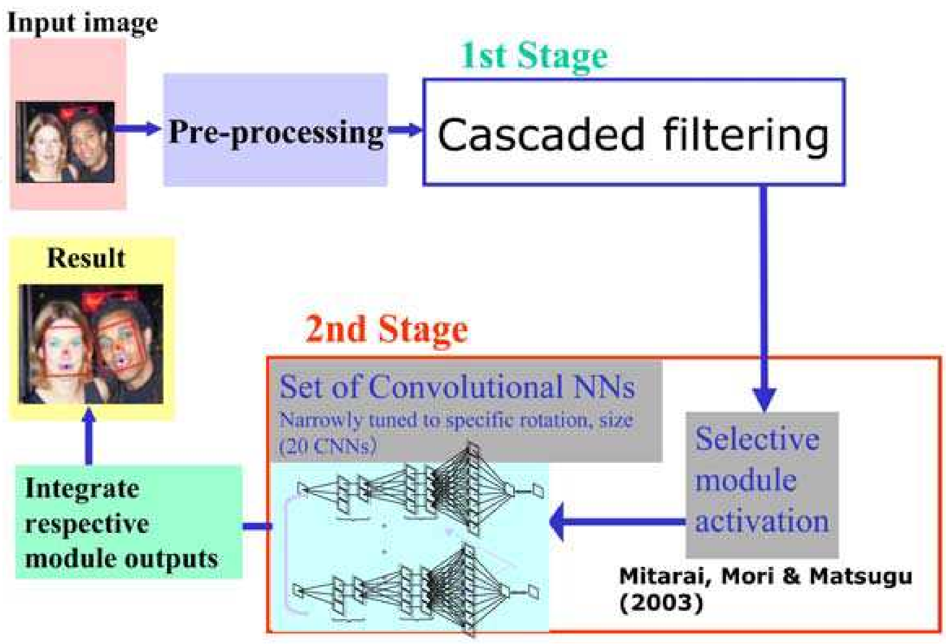


Fig. 3. Face detection process as a whole in the active vision system

3.2 Bottom-up process: integral feature description and matching using intersection

The bottom-up process involves extraction of local color cues inside attended region. In the proposed system, we used local color histogram as feature descriptor, since it potentially gives stable representation against noise without losing spatial information. We introduce a variant of integral representation of such features in order for fast computation of tracking features necessary for agility both in tracking (Porikli, 2005; Matsugu, et al., 2006) and in attention control (Frintrop et al., 2007). The integral representation of color histogram is given in (1).

$$H_{Y,I,Q}(x,y) = \sum_{x' \le x,\ y' \le y} b_{Y,I,Q}(x',y') \tag{1}$$

, where H(x, y) is the integral data at (x, y). Y, I, and Q are luminance and chrominance color coordinate value. The value of $b_{Y',I',Q'}$(x, y) is 1, if the bin with representative value of (Y', I', Q') is not empty at (x, y), and otherwise, $b_{Y',I',Q'}$(x, y) is zero. The local color histogram in a rectangle region can be computed with reference to integral data at four corner points of the region (Crow, 1984). Face location can be predicted using local histogram matching by using difference measure. In this paper, we use histogram intersection (Swain & Ballard, 1991; Birchfield & Rangarajan, 2005) between local image patch *I(R)* and *I(R')*, which is defined by (2).

$$\phi(I(R),I(R')) = \frac{\sum_{Y,I,Q} \min\left(\sum_{R} H_{Y,I,Q}(R), \sum_{R'} H_{Y,I,Q}(R')\right)}{\sum_{Y,I,Q}\sum_{R} H_{Y,I,Q}(R)} \tag{2}$$

, where R and R' are rectangle regions for local color histogram, and $H_{Y,I,Q}(R)$ is the area-summation (Crow, 1984) of local color histogram by using four H values given by (1) at respective corner points of the rectangle region R. Although the histogram intersection is not scale invariant, it is simple and effective for reliability and quick response of tracking even when the object moves with drastic appearance change in size (e.g., face, approaching the camera with much higher speed).

3.3 Feature coherency and autonomous switching between detection and tracking

The proposed system, without saliency map, *consolidates* respective streams of predicted object locations from bottom-up process by histogram matching (HM) and those from top-down process by face detection (FD). In the system, attention is covert, and attention shift is activated in an event driven manner so that tracking is maintained as long as possible. Once the system shall lose the sight of the tracked object, the top-down process (FD) dominates to search for the 'lost' object, and tracking is maintained by the bottom-up process (HM). Covert attention shift is activated when the lost object is detected inside the attention window in a way that attention window as well as camera pan/tilt angle is shifted to the predicted location (angle) of the object obtained from the FD process.

Control of pan/tilt camera head (details are given in Section 4) is accompanied by autonomous switching between two operation modes; detection mode and tracking mode. The duality and switching in between these two modes has also been explored in the literature (Chesi et al., 2003; Morioka et al., 2006; Yang et al., 2006). Distinct aspect of the present scheme as compared with similar past works (Sigal et al., 2004; Li et al., 2007) is the autonomous switching that results from integration of detection and tracking through a measure of coherence of top-down and bottom-up features. In view of preceding and current positional data (i.e., location queue of attention window), q, used for actual control, the coherency is given in (3).

$$C(p_{HM}, p_{FD}, q) = \min\left(dist(p_{HM}, q), dist(p_{FD}, q)\right) \quad (3)$$

, where $dist(p_s, q)$ is the distance between $\boldsymbol{p}_s$ (predicted position vector of object obtained from s process: FD or HM) and $\boldsymbol{q}$ (actual position vectors of preceding attention window). The coherency serves as tolerance parameter indicating whether or not hypothesized (predicted) location data are in accordance with previous control data. If p_{FD} for the current frame is not available due to occlusion or failure of object detection, p_{FD} is given as some fixed large value, C_0 (e.g., frame size) such that for any coherency value C, $C_0 > C$.

By using the above coherency measure, the consolidated tracking cue, $F(t)$, for the frame number t in Fig.1 for the pan/tilt control is drawn as follows.

$$F(t) = \begin{cases} p_{FD}(t-1) & \text{if top-down signal, } p_{FD}, \text{ available and } C = dist\ (p_{FD}, F(t-1)) \\ p_{HM}(t-1) & \text{if } C = dist\ (p_{HM}, F(t-1)) \text{ or } |p_{HM}(t) - pred(F(t))| < |p_{HM}(t) - F(t-1)| \\ F(t-1) & \text{otherwise} \end{cases} \quad (4)$$

, where $pred(F)$ is the linear prediction obtained from the preceding data F for more than 250ms. For faster calculation, we will simply use absolute difference for $dist(\boldsymbol{p}, \boldsymbol{q}) = |\boldsymbol{p} - \boldsymbol{q}|$.

3.4 Adaptive search area

Size and location of attention window are updated based on averaged prediction error for a fixed time interval (e.g., 150ms). This adaptive scheme is effective for increasing stability of tracking and to facilitate recovery from face detection failure and prediction error. The prediction error is defined by distances in horizontal and vertical directions on image plane, between predicted face location and the image center when face is detected. In the case of missing object (face) to be tracked, the error is given by *arctan* of the last detected face size multiplied by a constant parameter. The size is gradually enlarged during the detection failure (missing the object to be tracked) inside the window. Fig. 4 illustrates how the attention window enlarges depending on the motion of the object and the difference between the center of window and object location. In the left side picture of Fig. 4, the attention window is enlarged in horizontal direction since the object is moving horizontally, whereas in the right picture, the window size is set larger due to missing (failure in the top-down process) of object in the previously set window.

Fig. 4. Variable attention window (rectangle region with pink line segments)

4. Hardware system

The prototype active vision system is composed of a small camera module, pan/tilt movement subsystem, DC motors, multi-rate hybrid control units. Most part of tracking control for each subsystem is implemented on FPGA board. Tracking and object detection algorithms run on PC as parallel tasks composed of multiple threads.
The best available instantaneous tracking speed was approximately 6 deg/40ms (pan) as shown in Fig. 5 (left). The result is equivalent to tracking a person running and switching motion direction approximately at 10m/sec with distance about 3m of the camera. By tuning control parameters to avoid overshooting, the result was obtained with estimated prediction data of face location given at every 10ms.
Fig. 6 shows the results of pan/tilt angle control and associated prediction, with the hybrid (feed-forward/feed-back) PD control system (details are given in Fig. 8). Linear prediction of face location based on observation sequence has resulted in noisy and unsteady sequence (pan and tilt prediction). Applying critical sensitivity analysis and appropriate tuning of gain parameters (i.e., P and D gain) of step response, the proposed hybrid control attained smoothness (stability) and agility of tracking as well. Observation period for prediction is over 250ms and more than three points of prior observations (e.g., predicted face locations or detected face locations) are utilized to obtain the linear prediction.

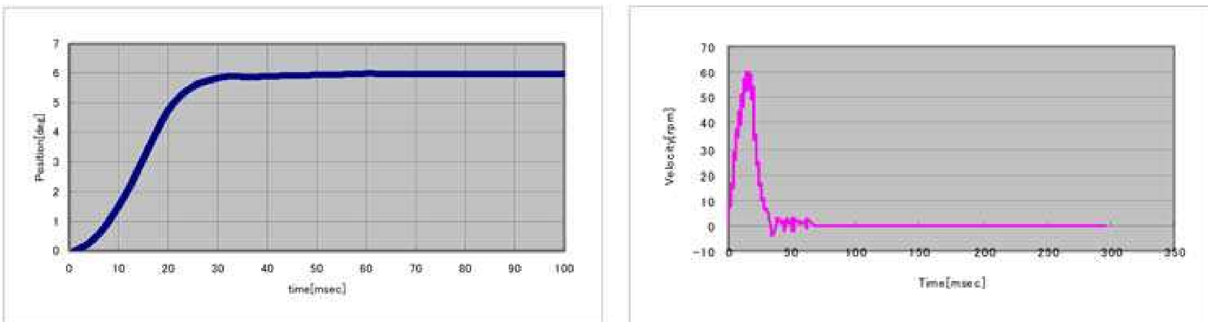


Fig. 5. Saccadic eyemovement without overshooting (left): angle-time, (right): vlocity-time

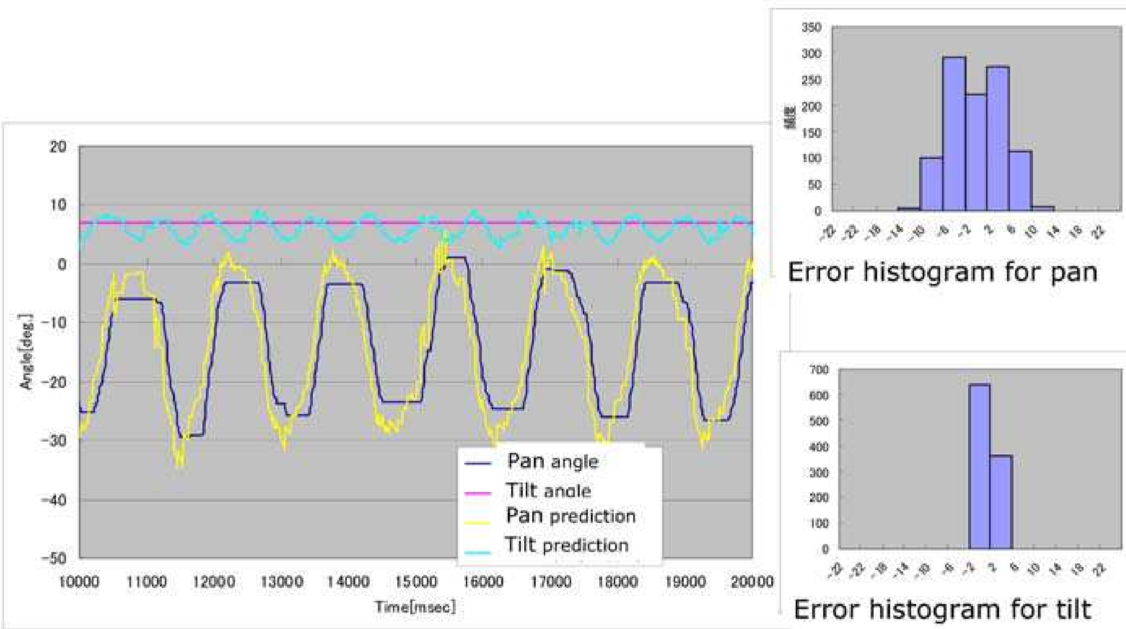


Fig. 6. Smooth and agile control with noisy prediction

The structure of pan/tilt system is a classic gimbal as shown in the middle and right side of Fig.7. Major component of the entire mechanical system with the time constant of 13ms are coreless DC motor (Maxon RE10), 1/40 reducer unit composed of three gears and a compact camera module (CK300: Keyence). Prototype picture of the vision head with display is given in the left of Fig. 7.

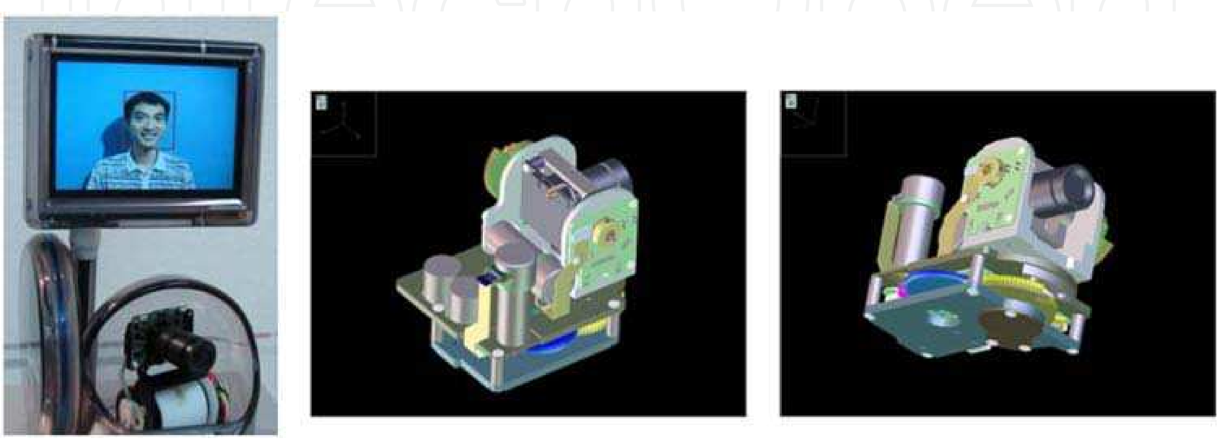


Fig. 7. Prototype active vision module with pan/tilt mechanical system

The maximum attainable pan/tilt speed is 900 deg/sec. The tilting mechanical unit is placed on top of panning unit, so that the panning has low inertia to realize quick response in horizontal direction. Thus, the tilting unit is accompanied by relatively larger inertia with gentler response as compared with the panning.
We employed a hybrid multi-rate control composed of FF based velocity control circuits and software operating at 100 Hz and FB based position control circuits operating at 5000Hz. The former helped realize smooth pursuit, and the latter contributed to the saccadic and quick response in our active tracking system. In Fig.8, most of components including camera controller and driver module, except for FF-velocity controller are implemented as hardware circuits on FPGA board (yellow block in Fig. 8).

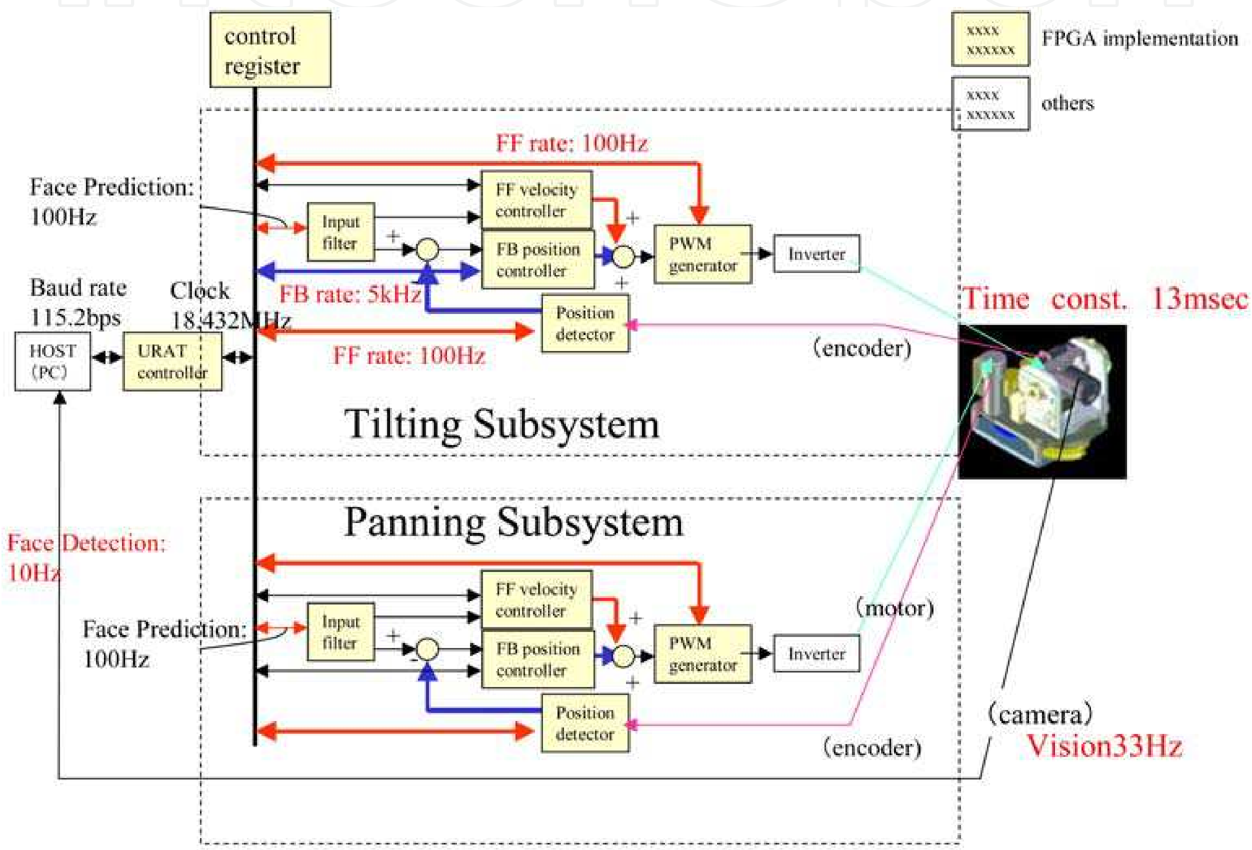


Fig. 8. Control system with fast feedback (5kHz) and slow feed-forward (100Hz) control signals

5. Results

Proposed system demonstrates remarkable response to track agile and nonlinear motion of object (person), as can be seen in Figs. 9, 10, and 11.
The minimum size of attention window is 3.7deg x 3.7deg and the maximum is 22.5deg x 22.5 deg. The system generated interpolated sequence of face locations for every 10ms. To suppress hopping from one subject to another during tracking in a scene where multiple faces present, maximum tracking speed of face motion is set as 0.1 deg/ms, leading to limitation of search window size.
Algorithms for detecting object to be tracked and prediction of the object location in our active vision system are implemented on PC (Pentium IV at 3.6 GHz). The system operates with five approximately concurrent processes: image capturing, object detection, object

prediction, camera control, and external output. As shown in Fig. 10, a person jumping with abrupt change in direction of motion can be tracked under the condition of heavy illumination change, and such tracking response can also be observed in panning sequence.

Fig. 9. Running sequence tracked with panning camera control.

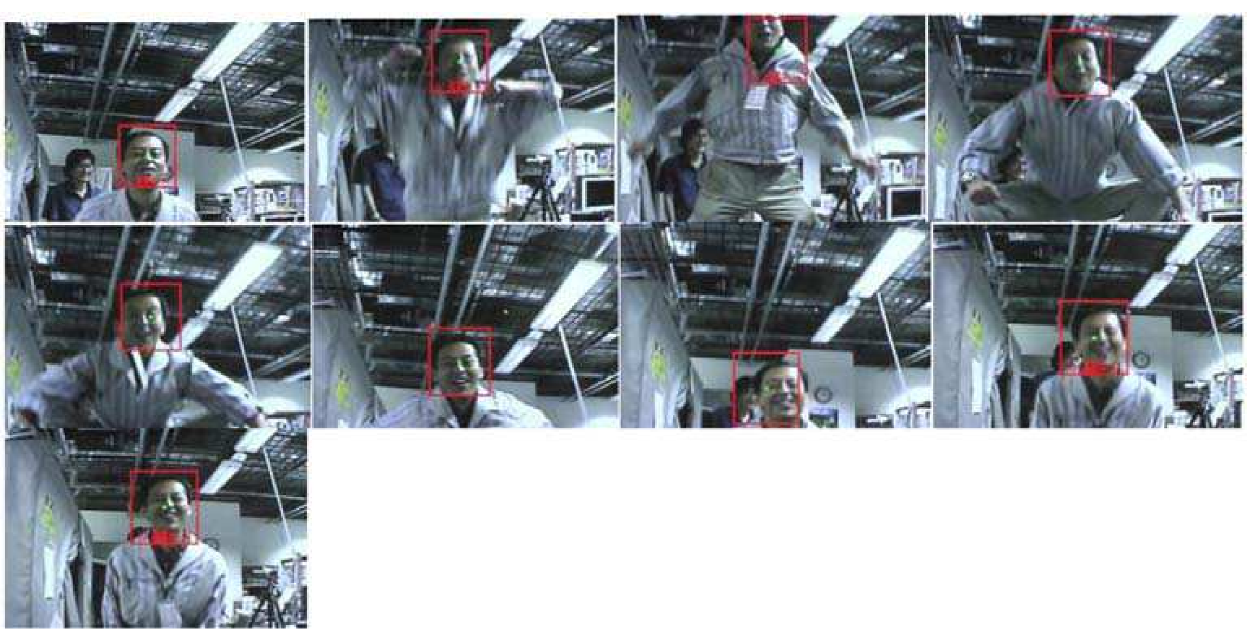


Fig. 10. Jumping sequence with motion blur, tracked with tilting camera control

Fig. 10 and Fig. 11 show tilting and panning control respectively for tracking sequences of jumping person with substantial motion blur. To further enhance the stability of tracking and to deal with temporary missing of tracked face due to occlusion or error in the top-down process, we introduced several heuristics; 1) inactivate pan/tilt control when located face is beyond the maximum speed of 0.1 deg/ms and no face location query is available in the time interval of 675ms, 2) use the last available object location obtained from the top-down process for pan/tilt control when the number of stored face location data that meet the coherency criterion (2) is less than four.
Fast recovery from error can be observed from these sequences. For example, the second picture in the top row of Fig. 11, the center of attention window is erroneous due to detection failure or prediction error, but in the next frame the center is correctly set. Such behavior results from the proposed attention control with feature consolidating mechanism proposed in Section 3, wherein both top-down and bottom-up processes are involved.

Fig. 11. Tracking a face of jumping person with rotation and motion blur

6. Discussion

Switching from detection (FD) mode to tracking mode with histogram matching (HM) is initiated typically when face detection is failed. Recovery from HM to FD mode is enforced when successful face detection inside the search window ensues. Stability in "visual servo" is thus realized by such covert attention shift and coherency based switching using top-down process. This event driven control turned out to be helpful for long-duration-tracking. Appropriate selection of control strategy is important to maximize performance (Papanikolopoulos et al., 1993), in terms of speed, stability, and robustness of active tracking systems. In this paper, we explored co-realization of saccadic and smooth pursuit using hybrid multi-rate control system (i.e., slow FF-velocity control and fast FB-position control). The proposed system integrated variety of modalities; fast processes (HM in bottom-up process and FB-position control) and slow processes (FD in top-down process and FF-velocity control). Such integrity of perception and control capable of long duration tracking

is a distinct aspect in contrast to similar active vision system (Castrillón-Santana et al., 1998). Using rather classical control strategy combined with the fast object detection and reliable computation of object cues from the image sequence, we successfully demonstrated quick and smooth response in visual servo and also effectiveness of proposed coherency based attention shift during tracking.
One notable point of this attention shift is fast recovery from failure (error) and another is the balance between quick response and smoothness in tracking. The former aspect is manly due to the covert attention shift that utilizes robust object detection, which is accompanied with autonomous switching between FD mode and HM mode. Integral image representation of local color histogram was indispensable for quick computation of attention cues, and adaptive setting of search area based on linear prediction errors was also effective for ensuring the top-down attention shift.
Even if faced with nonlinear motion and abrupt change of moving direction of the tracked object, no sopshisticated nonlinear control strategies are required due to covert attention shift resulting from top-down process, while preserving agility in response. This result was obtained partly because fast computation and selection of reliable bottom-up cues. The multi-rate (i.e. slow FF velocity control and fast FB position control) control turned out to be simple and computationally inexpensive, in contrast to existing methods (Yang et al., 2006; Morioka et al., 2006), and it could also strike a balance between saccade and smooth pursuit.

7. Conclusion

In this study, the task oriented active vision with top-down and bottom-up attention control demonstrated long duration tracking under bad illumination conditions, fast recovery from tracking error and quick response to abrupt change in moving direction of the tracked object. Nonlinearity of the object motion is handled by the proposed attention control without recourse to state-of-the art nonlinear control strategy. The consolidating mechanism between top-down and bottom-up features through coherency measure can serve as substrate for stable tracking of specific object with covert attention shift.

8. Acknowledgements

We thank T. Osaka (Fig. 7), T. Hayashi (Figs. 5, 6, and 8), and Onodera (Fig. 7) for the implementation and design of the proposed system.

9. References

Aloimonos, J. Weiss, I. & Bandyopadhyay, A. (1987). Active Vision. *International Journal of Computer Vision*, Vol. 1, No. 4, 333-356
Ballard, D. H. & Brown, C. M. (1992). Principles of Animate Vision. *Computer Vision, Graphics and Image Processing*, Vol. 56, 3-21
Birchfield, S. (1997). An elliptic head tracker. Proceedings of 31st Asilomar Conference on Signals, Systems, and Computer, pp. 1710-1714

Birchfield, S. & Rangarajan, S. (2005). Spatiograms versus histograms for region-based tracking, *Proceedings of IEEE International Conference on Computer Vision Pattern Recognition CVPR'05*, pp. 1158-1163

Blake, A. & Yuille, A. (1992). *Active Vision*, MIT Press, ISBN: 978-0262023511, Cambridge

Bradshaw, K. J., McLauchlan, P. F., Reid, I. D. & Murray D. W. (1994). Saccade and Pursuit on an Active Head/Eye Platform. *Image and Vision Computing,* Vol. 12, 155-163

Bur, A., Wurtz, A., Müri, R. M. & Hügli. H. (2007). Dynamic visual attention: competitive and motion priority scheme. *Workshop on Computational Attention & Application in International Conference on Computer Vision Systems*

Burt, P.J., Bergen, J.R., Hingorani, R. & Kolczynski, R. (1989). Object tracking with a moving camera. *Proceeding of Workshop on Visual Motion*

Castrillón-Santana, M., Guerra-Artal, C., Hernández-Sosa, J., Domínguez-Brito, A., Isern-González, J., Cabrera-Gámez, J. & Hernández-Tejera, F.M. (1998). An Active Vision System Integrating Fast and Slow Processes. *Proceeding of the SPIE'98 Symposium on Intelligent Systems and Advanced Manufacturing,* pp. 487-496, Boston

Chesi, G., Hashimoto, K., Prattichizo, D. & Vicino, A. (2003). A switching control law for keeping features in the field of view in eye-in-hand visual servoing. *Proeedings of IEEE International Conference on Robotics and Automation* (Taipei, Taiwan), pp. 3929-3934

Clark, J. J. & Ferrier, N. J. (1992). Attentive Visual Servoing. *Active Vision*, Blake, A. & Yuille, A. (Eds.), 137-154, MIT Press, ISBN: 978-0262023511, Cambridge

Comaniciu, D., Ramesh, V. & Meer, P. (2000). Real-time tracking of non-rigid objects using mean shift. *Proceedings of International Conference on Computer Vision & Pattern Recognition*, pp. 142-149

Corbetta, M. & Shulman, G. L. (2002). Control of Goal-direct and Stimulus-driven Attention in the Brain. *Nature Neuroscience*, Vol. 3, 201-215

Cretual, A., Chaumette, F. & Bouthemy, P. (1998). Complex object tracking by visual servoing based on 2D image motion. *Proceedings of International Conference on Pattern Recognition*, pp. 1251-1254

Crow, F. (1984). Summed-Area Tables for Texture Mapping. *Computer Graphics*, Vol. 18, 207-211

Culhane, S. M. & Tsotsos, J. K. (1992). A Prototype for Data-Driven Visual Attention. *Proceedings of International Conference on Pattern Recognition*, pp.36-40

Drummond, T. & Cipolla, R. (2002). Real-Time Visual Tracking of Complex Objects. *IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol. 24, No. 7, 932-946*

Frintrop, S., Klodt, M. & Rome, E. (2007). A Real-time Visual Attention System Using Integral Images, *Proceedings of International Conf. on Computer Vision Systems*

Fujita, M., Kawai, H. & Spong, M. W. (2007). Passivity-based dynamic visual feedback control for three dimensional target tracking: stability and L2-gain performance analysis. *IEEE Transactions on Control Systems Technology*, Vol. 15

Garvey, T. D. (1976). Perceptual strategies for purposive vision. *Technical Note* Vol. 117, SRI International

Giefing, G.-J., Janβen, H. & Mallot, H. (1992). Saccadic Object Recognition with an Active Vision System. *Proceedings of European Conference on Artificial Intelligence*, pp. 803-805

Hopfinger, J. B., Buonocore, M. H. & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, Vol. 3, 284-291

Hunt A. E. & Sanderson, A. C. (1982). Vision-based predictive robotic tracking of a moving targets. *Technical Report*, Carnegie Mellon University,

Isard, M. & Blake, A. (1998). ICondensation: Unifying Low-level and High-level Tracking in a Stochastic Framework. *Proceedings of European Conference on Computer Vision*, pp. 893-908

Itti, L., Koch, C. & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 20, 1254-1259

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Research*, Vol. 40, 1489-1506

Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: Towards the under-lying neural circuitry. *Human Neurobiology*, Vol. 4, 19-227

Kyrki, V. & Kragic, D. (2005). Integration of Model-based and Model-free Cues for Visual Object Tracking in 3D. *Proceedings of International Conference on Robotics and Automation*

Lee, K. W., Buxton, H. & Feng, J. (2005). Cue-Guided Search: A Computational Model of Selective Attention. *IEEE Transactions on Neural Networks*, Vol. 16, 910-924

Li, Y., Ai, H., Huang, C., & Lao, S. (2006). Robust head tracking based on a multi-state particle filter. *Proceedings of Automatic Face & Gesture Recognition,* pp.335-340

Li, Y., Ai, H., Yamashita, T., Lao, S. & Kawade, M. (2007). Tracking in Low Frame Rate Video; A Cascade Particle Filter with Discriminative Observers of Different Lifespans. *Proceedings of International Conference on Computer Vision & Pattern Recognition*.

Mancas, M.; Gosselin, B. & Macq, B. (2007). A Three-Level Computational Attention Model. *Workshop on Computational Attention & Application in International Conference on Computer Vision Systems.*

Matsugu, M. & Cardon, P. (2004). Unsupervised feature selection for multi-class object detection using convolutional neural networks. *Advances in Neural Networks – ISNN 2004 International Symposium on Neural Networks, LNCS 3173,* 864-869, Springer, ISBN: 978-3-540-22841-7, Berlin

Matsugu, M., Mori, K. & Mitarai, Y. (2004). Convolutional Spiking Neural Network for Robust Object Detection with Population Code Using Structured Pulse Packets. *Neural Information Processing: Research and Development*, Rajapakse, J.C. & Wang, L. (Eds.), 39-55, Springer, ISBN: 3-540-21123-3, Berlin

Matsugu, M., Torii, K., Ito, Y., Hayashi, T. & Osaka, T. (2006). Face Tracking Active Vision System with Saccadic and Smooth Pursuit. *Proceedings of IEEE International Conference on Robotics and Biomimetics.*

Mitarai, Y., Mori, K. & Matsugu, M. (2003). Robust face detection system based on convolutional neural networks using selective activation of modules (In Japanese), *Proceedings of Forum in Information Technology*, pp. 191-193

Morioka, K., Kuroda, Y., Lee, J.-H. & Hashimoto, H. (2006). Multiple objects tracking based on auto-selection of color image processing methods. (in Japanese) *IEEJ Trans. EIS*, Vol. 126, No.2, 210-219

Murao, T. , Yamada, T. & Fujita, M. (2006). Predictive visual feedback control with eye-in-hand system via stabilizing receding horizon approach. *Proceedings of the 45th IEEE Conference on Decision and Control,* pp. 1758-1763

Murray, D. & Basu, A. (1994). Motion tracking with active camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 16, 449-459

Navalpakkam, V. & Itti, L. (2006). Top-down attention selection is fine grained. *Journal of Vision*, Vol. 6, 1180-1193, ISSN: 1534-7362

Olshausen, B.; Anderson, C.H. & Van Essen, D.C. (1995). A Multiscale Dynamic Routing Circuit for Forming Size- and Position-Invariant Object Representations. *Journal of Computational Neuroscience*, Vol. 2, 45-62, ISSN: 0929-5313

Papanikolopoulos, N. P., Khosla, P. K., & Kanade, T. (1993). Visual tracking of a moving object by a camera mounted on a robot. *IEEE Transactions on Robotics and Automation*, Vol. 9, 14-35

Porikli, F. (2005). Integral histogram: a fast way to extract histograms in Cartesian spaces. *Proceedings of International Conference on Computer Vision ond Pattern Recognition,* pp. 829 - 836

Porikli, F., Tuzel, O. & Meer, P. (2006) Covariance Tracking Using Model Update Based on Lie Algebra. *Proceedings of International Conference on Computer Vision ond Pattern Recognition*.

Rasmussen, C. & Hager, G. D. (2001). Probabilistic Data Association Methods for Tracking Complex Visual Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 23, 560-576

Ribraic, S., Adrineck, G. & Segvic, S. (2004). Real-time active visual tracking system. *Proceedings of IEEE Mediterranean Electrotechnical Conference on Pattern Recognition,* pp. 231-234

Sigal, L., Zhu, Y., Comaniciu, D. & Black, M. (2004). Tracking Complex Objects Using Graphical Object Models. *Proceedings of 1st International Workshop on Complex Motion*, Springer- Verlag, Vol. LNCS 3417, pp. 227-238

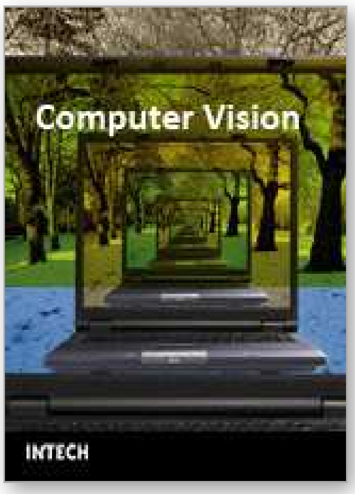
Serby, D., Koller-Meier, E. & Van Gool, L. (2004). Probabilistic Object Tracking Using Multiple Features. *Proceedings of International Conference on Pattern Recognition,* pp. 184-187

Sun, Y. & Fisher, R. (2003). Object-based Visual Attentions for Computer Vision. *Artificial Intelligence*, Vol. 146, 77-123

Swain, M.J. & Ballard, D.H. (1991). Color Indexing. *International Journal of Computer Vision,* Vol. 7, 11-32

Swain, M.J., Kahn, R. & Ballard, D.H. (1992). Low Resolution Cues for Guiding Saccadic Eye Movements. *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 737-740

Viola, P. & Jones, M. (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features. *Proceedings of IEEE Computer Vision on Pattern Recognition*
Yang, S., Li, S. Z., Pan, Q., Li, J. & Zhao, C. (2006). Reliable and fast tracking of faces under varying pose. *Proceedings of Automatic Face and Gesture Recognition,* pp.421-426
Yang, M., Yuan, J. & Wu, Y. (2007). Spatial selection for attentional visual tracking. *Proceedings of International Conference on Computer Vision and Pattern Recognition*



Computer Vision
Edited by Xiong Zhihui

ISBN 978-953-7619-21-3
Hard cover, 538 pages
Publisher InTech
Published online 01, November, 2008
Published in print edition November, 2008

This book presents research trends on computer vision, especially on application of robotics, and on advanced approachs for computer vision (such as omnidirectional vision). Among them, research on RFID technology integrating stereo vision to localize an indoor mobile robot is included in this book. Besides, this book includes many research on omnidirectional vision, and the combination of omnidirectional vision with robotics. This book features representative work on the computer vision, and it puts more focus on robotics vision and omnidirectioal vision. The intended audience is anyone who wishes to become familiar with the latest research work on computer vision, especially its applications on robots. The contents of this book allow the reader to know more technical aspects and applications of computer vision. Researchers and instructors will benefit from this book.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821