Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Adaptive Visual Tracking for Surveillance Systems ⋆

Seok Min Yun * **Jin Hee Na** * **Jin Young Choi** *
Myung Seok An ** **Myung Ho Yoo** **

* *Electrical Engineering Department, Automation and System Research Institute(ASRI), Seoul National University, Seoul, Korea, (e-mail: {smyoon, jhna}@neuro.snu.ac.kr, jychoi@snu.ac.kr)*
** *Samsung Techwin Co.,LTD, (e-mail: {myungseok.an, myungho.yoo}@samsumg.com)*

Abstract: In this paper, we propose a new hierarchical estimation approach for adaptive visual tracking. This approach includes incremental appearance model and hierarchical estimation: global estimation and local estimation. Local estimation is performed by the particle filter for dealing with non-linearities and non-Gaussian statistics, and global estimation is performed by Kalman filter to determine the scatter positions of particles for local estimation. By combining these two estimations, the number of particles used in local estimation can be reduced in global estimation, and it enables real-time tracking while maintaining or improving tracking abilities. Experimental results show the effectiveness and robustness of the proposed approach compared with those of existing tracking method.

1. INTRODUCTION

Visual tracking is essential feature in surveillance systems as well as many computer vision applications. The implementation of visual tracking against real world is challenging problem due to intrinsic and extrinsic variations. Intrinsic variation include appearance change, shape deformation and pose variation of target object while extrinsic variantion include illumination variation, camera motion, dense and dynamic background clutter and occulusions. In tracking algorithm, these variations inevitably result in large unforeseen appearance changes which is the principal cause of failure.

To overcome these problems, several kinds of tracking algorithms have been proposed. Most of them can be classified into two approaches: deterministic tracking and stochastic tracking(Shaohua kevin Zhou et al. (2004)). Deterministic tracking approaches usually reduce to an opmization problem. Typically they perfom iterative search to minimize appropriate cost function. Based on the definition of the cost function many tracking methods have been derived. Model-based tracking, Appearance-based tracking and Mean-shift(D. Comaniciu et al. (2000)), EigenTracking(M. J. Black et al. (1996)) are exactly the case. The EigenTracking approach essentially attempted to establish a robust appearance model based on view-based eigenbasis representation. Although their algorithm demonstrated excellent empirical results and shows its robustness in some appearance change, while it needs pre-training process in advance before tracking and furthermore its robustness is only restricted in trained apperance variations which do not support on-line updating.

On the other hand, stochastic tracking approaches often consider as estimation problem, e.g., estimating the state space to model the underlying dynamics of the tracking system. In early work, Kalman filter and its variants are used to provide solutions. However, this method restricted the model in Linear Gaussian case. For non-linear or non-Gaussian problems, the particle filter, also known as sequential Monte Carlo algorithm(A. Doucet et al. (2001)) have gained prevalence in the tracking literature due to the Condensation algorithm(M. Isard et al. (1996)). It is one powerful methodology for maintaining non-Gaussian distributions. While this algorithm is also vulnerable to appearance or background change due to its fixed representation of target object.

Recently, incremental or adaptive representation which can model varying appearance manifold has become novel approach. Lim et al. (2005) proposed incremental method and Shaohua kevin Zhou et al. (2004) proposed adaptive method. Both approaches adopt the integration of deterministic and stochastic tracking methods. For deterministic method, they use incremental or adaptive appearance model instead of fixed appearance model and both of them utilize particle filter in stochastic method. Lim et al. (2005) adopt Gaussian dynamic model in his particle filter, which is vulnerable to nonlinear motion. As the consequence, it needs much more particles. Shaohua kevin Zhou et al. (2004) proposed adaptive velocity motion model, where the adaptive motion velocity is predicted using a first-order linear approximation based on apprearance change. In addition, they also use adaptive number of particles in particle filter which makes it works more efficiently. Though their adaptive velocity motion model is more powerful than linear Gaussian dynamic model in handling nonlinear motion, it uses first-order linear approximation which is incapable of measuring acceleration term.

This paper addresses a solution to nonlinear motion in visual tracking. We propose a hierarchical estimation approach which is consist of global estimation and local es-

⋆ This research was supported by Ministry of Commerce, Industry and Energy of Korea and Samsung Techwin Co., Ltd.

timation. Global estimation is performed by Kalman filter to determine the initial position for potential particles, while local estimation is in charge of dealing with nonlinearities and non-Gaussian statistics by using particle filter. By hiearchically combining these two estimations, our algorithm not only can significantly reduce particles without loss of accuracy performance, but also gives real-time tracking capability with improving its robustness. This strategy is based on the phenomenon that most of natural motion fit linear Gaussian model in global view, global scale and nonlinear motion is usually restricted in local view.

We have tested our algorithm on video sequences of human faces with non-linear rapid moving. Experimental results show the effectiveness of the proposed approach compared with those of existing tracking methods.

This paper is organized as follows. We briefly review the related literature on incremental appearance model and particle filter in Section 2. We show the details of hierarchical estimation approach in Section 3, and experimental results on several examples in Section 4. At last conclusions are presented in Section 5.

2. PROBLEM STATEMENT

State-space approach in visual tracking requires estimation of system state that changes over time using a observation sequence(Arulampalam et al. (2002)), and usually consists of dynamic model(or system model) and observation model (or measurement model). In this approach, we can assume that these models are available in a probabilistic form that is suited for the Bayesian tracking. Then, visual tracking problem is replaced as an inference problem with a Markov model and hidden state variable, where X_t is a set of state variables at time t, and $\mathcal{Z}_t = \{Z_1, \cdots, Z_t\}$ is a observation set. If a observation Z_t is given, we have posterior probability as follows using Bayes' rule.

$$p(X_t|\mathcal{Z}_t) = \frac{p(Z_t|X_t)p(X_t|\mathcal{Z}_{t-1})}{p(Z_t|\mathcal{Z}_{t-1})}. \tag{1}$$

Applying Chapman-Kolmogorov equation to $p(X_t|\mathcal{Z}_{t-1})$, sequential inference model is given as follows.

$$p(X_t|\mathcal{Z}_t) \propto p(Z_t|X_t) \int p(X_t|X_{t-1})p(X_{t-1}|\mathcal{Z}_{t-1})dX_{t-1}. \tag{2}$$

Note that tracking process depends on the observation model $p(Z_t|X_t)$ and dynamic model $p(X_t|X_{t-1})$, and a particle filter is famous method to obtain the required posterior probability using a set of samples(particles) with propagating sample distribution.

For observation model, we represent observations using incremental subspace update method suggested in Lim et al. (2005). This method is based on R-SVD methods(G. H. Golub and C. F. Van Loan (1996)), and update the eigenbasis in on-line phase while updating the sample mean into account. The generation number of particles in subspace is inversely proportional to the distance from the subspace center(i.e., sample mean) to the particles, and this distance is decomposed into the distance-to-subspace,d_t, and the distance-within-subspace d_w. The likelihood of the particles, $p(Z_t|X_t)$, are given by a Gaussian distribution.

$$p(Z_t|X_t) = p_{d_t}(Z_t|X_t)p_{d_w}(Z_t|X_t), \tag{3}$$

where $p_{d_t}(Z_t|X_t)$ and $p_{d_w}(Z_t|X_t)$ are also Gaussian. Details to compute d_t, d_w and the likelihood of particle are described in Lim et al. (2005).

To the best of our knowledge, the probability of state transition is defined in previously, and the each state variable of X_t is diffused independently from that of X_{t-1} according to Gaussian distribution. That is,

$$p(X_t|X_{t-1}) = \mathcal{N}(X_t; X_{t-1}, \Psi). \tag{4}$$

Where, ψ is a diagonal matrix that includes variances of state variables. In equation (4), each element of X_{t-1} is considered as mean position of diffusion. Namely, particles are generated around the elements of X_{t-1} according to Gaussian distribution in probability sense, and estimation error in the previous step can affect the estimation in the next step. Increasing the number of particles can be the method to reduce the estimation error rate, however it requires large amounts of computation time. Considering this fact, we modify the dynamic model as a hierarchical structure: global estimation and local estimation. Details are described in the next section.

3. HIERARCHICAL ESTIMATION FOR VISUAL TRACKING

3.1 Concept of Hierarchical Estimation

A Kalman filter models the state of a system using a Gaussian probability density function which propagates over time(Gong et al. (2000)), and is paramiterized by a mean and covariance. In this work, we use a second-order Kalman predictor with $\mu = (\mu_{x,t}, \dot{\mu}_{x,t}, \ddot{\mu}_{x,t}, \mu_{y,t}, \dot{\mu}_{y,t}, \ddot{\mu}_{y,t})^T$, and it is applied for determining the initial particle positions of local estimation. We model the state of global estimation as follows.

$$\mu_t = A\mu_{t-1} + w_{k-1}, \tag{5}$$

where, A is the block-diagonal matrix:

$$A = \begin{pmatrix} B & 0 \\ 0 & B \end{pmatrix} \quad B = \begin{pmatrix} 1 & \triangle t & \triangle t^2/2 \\ 0 & 1 & \triangle t \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$

with a measurement z_k that is,

$$z_k = Hx_k + v_k, \tag{7}$$

where,

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \tag{8}$$

The random variables w_{k-1} and v_k represent the process and measurement noise respectively. Here, we assume that w_{k-1} and v_k have zero mean and are statistical independent of each other with normal distributions

$$p(w) \propto \mathcal{N}(0, Q), \tag{9}$$

$$p(v) \propto \mathcal{N}(0, R), \tag{10}$$

where Q is process noise covariance and R is measurement noise corvariance. Here, we assume they are constant. Details for Kalman filters describe in Greg Welch and Gary Bishop (2001) . For local estimation, we use the six variables of affine transform to model the state evolution from X_{t-1} to X_t. X_t is defined like this: $X_t = (x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t)^T$, where x_t, y_t, θ_t, s_t, α_t, ϕ_t denotes x position, y position, rotation, angle, scale, aspect ratio and skew direction respectively.

Local estimation begins with the result of Kalman filter.

As described in equation (4), the likelihood of particle, $p(Z_t|X_t)$, is determined by two distance based probabilities $p_{d_t}(Z_t|X_t)$ and $p_{d_w}(Z_t|X_t)$. Given an predicted image patch, X_t, the likelihood of a particle generated in incremental subspace is parameterized by the result of Kalman prediction(mean) with variance terms defined in Lim et al. (2005).

$$\begin{aligned} p(Z_t|X_t) =& p_{d_t}(Z_t|X_t)p_{d_w}(Z_t|X_t) \\ =& \mathcal{N}(Z_t;\mu,UU^T+\varepsilon I)\mathcal{N}(Z_t;\mu,U\Sigma^{-2}U^T). \end{aligned} \tag{11}$$

where the columns of U are eigenvectors, and Σ is the diagonal matrix of singular values corresponding to the column of U.

3.2 Effect of Hierarchical Estimation

When applying conventional particle filter for adaptive visual tracking, the posterior density function is represented by a set of random particles in the incrementally updating subspace. The particles with high weights have been duplicated many times with high probabilities in the resampling step, and tracking point is determined as the maximum probability particle using equation (4). As shown in Lim et al. (2005), this approach works well even though the target objects undergo pose and lighting changes by adaptation of model representation. However, the errors in the estimation part of tracking system can make the system adapt to inappropriate targets. Figure 1 shows the tracking failure which is caused by errors in particle filter estimation. When the object moves too fast to tracking, particles can be scattered at the remoted region from ground truth data('□'). Increasing the number of particles makes it possible that maximum probability particle is located near ground truth data, however, it degradates the tracking efficiency.
In this work, we reduce the estimation errors using hierarchical estimation. Figure 2 shows the tracking result based on the hierarchical estimation using Kalman filter and particle filter. First, the mean scatter position of particles is predicted by Kalman filter which makes the particles are scattered nearby real tracking position, and the particle which has maximum probability is determined as a final estimation result('◇') at every frame. This is a kind of '*coarse-to-fine*' search that consists of global estimation(Kalman filter) and local estimation(particle filter). By including global estimation step, we can use the reduced number of particles and improve the tracking accuracy and efficiency.

4. EXPERIMENTAL RESULTS

4.1 Experimental Setup

For comparison with previous tracking approach, we performed experiments on the public available sequences in which target moves with several variations, such as occlusion, pose and illumination changes. In addition, we also performed experiment to demonstrate the property of our tracking approach. All images are gray-scale images, and the tracking area has been initialized manually in the first frame. Each image patch for tracking is resized to 32×32 for adaptive subspace update. We computed the Mean Square Error(MSE) between the estimated tracking

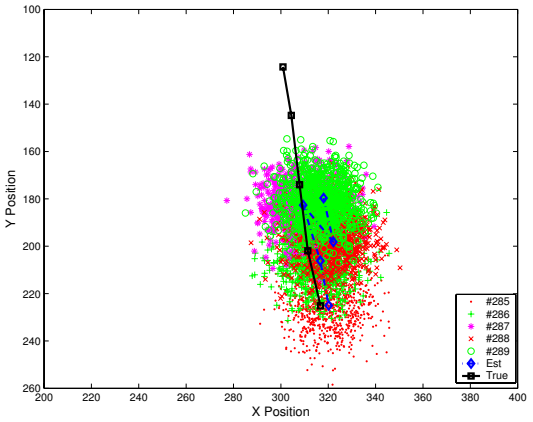

Fig. 1. Result of particle filter estimation. The propagation of particles in 5 sequences(♯285–♯289 frames) are represented with ground truth data indicated by '□'. The estimation result is indicated by '◇'. 800 particles are used at each propagation.

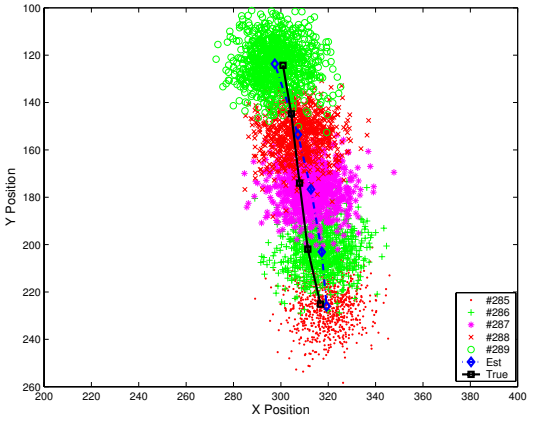

Fig. 2. Result of hierarchical estimation. The propagation of particles in 5 sequences(♯285–♯289 frames) are represented with ground truth data indicated by '□'. The estimation result is indicated by '◇', and the scatter positions of particles are more suitalble for estimation. 800 particles are used at each propagation.

results and the ground truth data to verify the tracking performance. All experiments have been performed on a standard 3.0GHz PC with 1GB RAM using MATLAB.

4.2 Evaluation Comparison

Figure 3 shows some image frames in the Dudek sequence, which is originally appeared in A. D. Jepson et al. (2001). This sequence contains a moving face in front of cluttered background, and it contains lots of activities which cause appearance changes, such as a hand occluding the face for a short time, taking the glasses on and off, and standing up rapidly, etc.. Using this sequence, we have compared the performance of our approach with that of other approach. Figure 4 illustrates the tracking failure of adaptive visual tracking with particle filter(Lim et al. (2005)) when standing up rapidly. The mean scatter position of particles are determined as the previous tracking position, and it is inadequate for tracking further even if using 800 particles. While the proposed approach with hierarchical

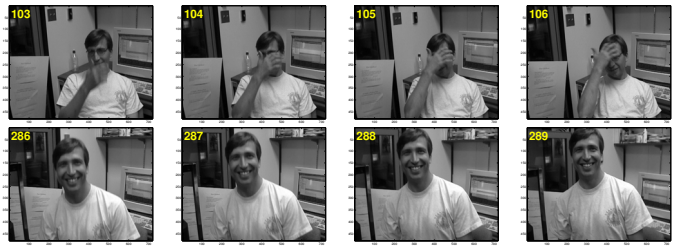

Fig. 3. Image sequence for face tracking. A face moves in front of cluttered background and contains variantions in appearance. This sequence has provided by Lim et al. (2005), and its original sequence is provided by A. D. Jepson et al. (2001). The first row illustrates a hand accluding the face for a short time, and the second row illustrates standing up rapidly. In addition, there are other appearance changes, such as glasses on and off, etc..

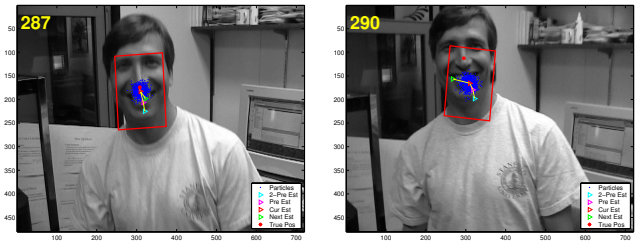

Fig. 4. Tracking results using particle filter with adaptive subspace update by Lim et al. (2005). When a face moves fast to the upper direction, tracking failure occurs. In this experiment, we use 800 particles in every frames, and achieve 4 frames/sec on our 3.0GHz PC.

estimation can track the face only using 100 particles in the same sequences, as shown in Figure 5. Note that our approach achieves better tracking performance while using the less number of particles. Our current implementation runs at 20 frames/sec with 100 particles, and 4 frames/sec with 800 particles without any code optimization. Figure 6 and Figure 7 illustrate the result comparison of MSE estimation. Our hierarchical estimation performs better than the previous particle filter based estimation in most frames without concerning the number of particles. The MSE results are given by as follows.

$$MSE(t) = \frac{1}{n_{gt}} \sum_{1}^{n_{gt}} (X_{est}(t) - X_{true}(t))^2 \qquad (12)$$

where, n_{gt} is the number of ground truth data in a sequence, $X_{est}(t)$ is the estimated position and $X_{true}(t)$ is the true position at time t. We used 7 ground truth data for each sequence given by Lim et al. (2005).

Figure 8 and Figure 9 show the tracking results of our approach and previous approach using another public available sequence, respectively. This sequence contains drastic illumination change combined with pose variantion. Only using particle filter with subspace update model, we can track a face until about 280 frames. However, when a face moves fast with illumination change, tracking failure occurs. While, our tracker successfully follows the face robustly at that fast moving frames. Note that our hierarchical approach is useful in a case like that; a target moves

Fig. 5. Tracking results of our approach. In this experiment, we use 100 particles in every frames, and achieve 20 frames/sec on our 3.0GHz PC.

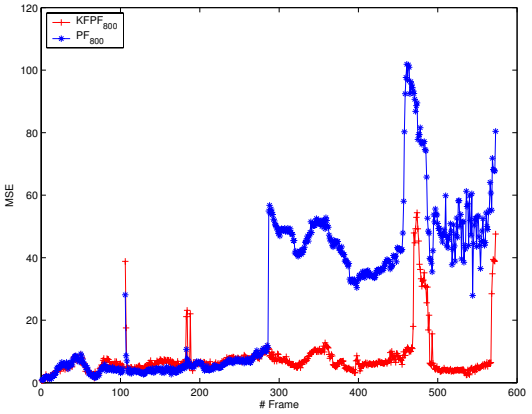

Fig. 6. Result comparison of MSE estimation using dudek sequence. In these experiments, 800 particles are used at every frames.

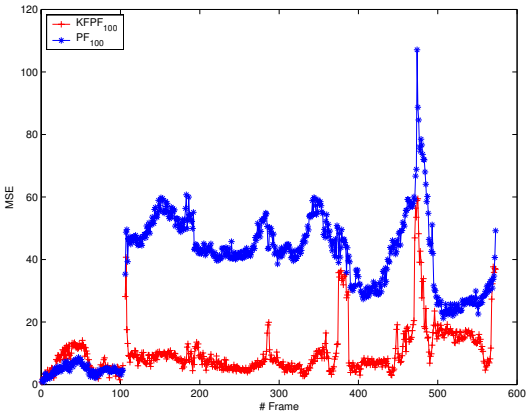

Fig. 7. Result comparison of MSE estimation using dudek sequence. In these experiments, 100 particles are used at every frames.

Fig. 8. Tracking results of our approach. This video sequence has been provided by Lim et al. (2005), and shows that a person moves underneath a trellis with large illumination changes. In their experiments, 600 particles are adopted, while we use 150 particles in our experiment.

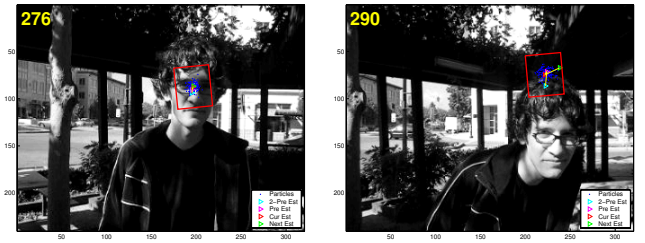

Fig. 9. Tracking results of Lim's original approach with 150 particles.

fast with other appearance changes. In these experiments, we use 150 particles in both cases.

Final experiments are shown in Figure 10. This result also shows that our tracker follows successfully with the appearance changes in pose ,illumination and partial occlusion. We use 150 particles in this experiment.

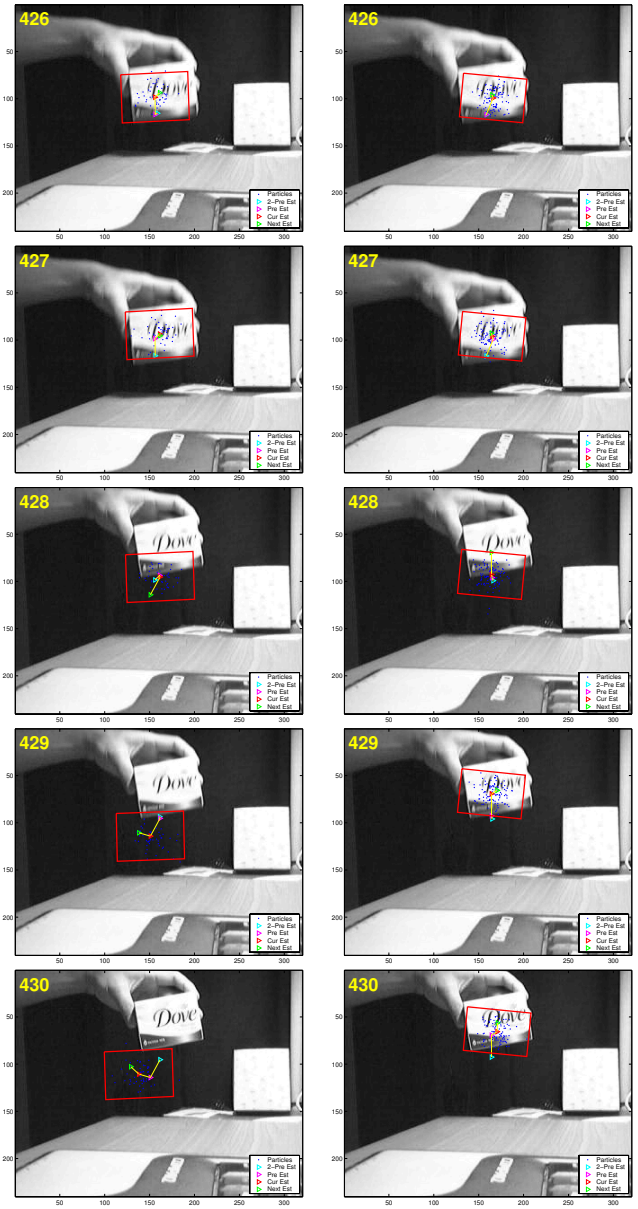

Fig. 10. Comparison of Tracking results. This video sequence shows a small box moves quickly and randomly. Left frames are the tracking results of Lim's approach with 50 particles, and right frames are those of our approach with same number of particles.

5. CONCLUSION

The accuracy and efficiency of the particle filter estimation depend on the particle propagation function for particle allocation and the number of particles. Considering this fact, in this paper, we proposed the hierarchical estimation using Kalman filter and particle filter for adaptive visual tracking. Particles are generated in Gaussian distribution which mean is determined by the Kalman estimation, and this method reduces the number of particles while maintaining or improving tracking abilities. Our approach is combined with incremental subspace update algorithm to adapt the variation of tracking region, that the center of subspace is determined by Kalman filter. We have shown the accurracy and the efficiency of the proposed

approach in several experiments using public available image sequences.
We expect that the proposed estimation method can be extended to the adaptive number of particles and pose variance(out of plane) estimation hierarchically similar with position estimation.

REFERENCES

Shaohua Kevin Zhou, rama Chellappa, and Baback Moghaddam. Visual Tracking and Recognition Using Appearance-Adaptive Models in Particle Filters. *IEEE Transations on Image Processing*,volume 13, No 11, pages, 1492–1493, November 2004.

D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of nonrigid objects using mean shift. *In Proceedings of IEEE Computer Society Conf. Computer Vision and Pattern Recognition*,volume 2, pages, 142–149, 2000.

M. J. Black, A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using view-based representation. *In Proceedings of European Conference on Computer Vision*, pages, 329–332, 1996.

M. Isard, A. Blake. Contour tracking by stochastic propagation of conditional density. *In Proceedings of Fourth European Conference on Computer Vision*, volume 2,pages, 343–356, 1996.

A. Doucet, N. d. Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.

M.Sanjeev Arulampalam, Simon Maskell, Neil Gordon and Tim Clapp. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing*, volume 50, pages 174–188, February 2002.

Jongwoo Lim, David Ross, Ruei-Sung Lin, Ming-Hsuan Yang. Incremental Learning for Visual Tracking. *Advances in Neural Information Processing Systems 17*, MIT Press, 2005.

G. H. Golub and C. F. Van Loan. *Matrix Computations*, The Johns Hopkins University Press, 1996.

Shaogang Gong, Stephen J. McKenna and Alexandra Psarrou. *Dynamic Vision: From Images to Face Recognition*, Imperial College Press, 2000.

Greg Welch and Gary Bishop. An Introduction to the Kalman Filter. *SIGGRAPH 2001 course pack edition*, ACM Press, 2001.

A. D. Jepson, D. J. Fleet and T. F. El-Maraghi. Robust Online Appearance Models for Visual Tracking. *In Proc. CVPR*, volume 1, pages 415–422, February 2001.