Probabilistic Index Histogram for Robust Object Tracking

Wei Li[1], Xiaoqin Zhang[2], Nianhua Xie[1], Weiming Hu[1], Wenhan Luo[1], Haibin Ling[3]

[1]National Lab of Pattern Recognition, Institute of Automation,CAS, Beijing, China.
{weili, nhxie, wmhu, whluo} @nlpr.ia.ac.cn
[2]College of Mathematics & Information Science, Wenzhou University, Zhejiang, China
[3]Dept. of Computer and Information Sciences Temple University
[2]xqzhang@wzu.edu.cn [3]hbling@temple.edu

Abstract. Color histograms are widely used for visual tracking due to their robustness against object deformations. However, traditional histogram representation often suffers from problems of partial occlusion, background cluttering and other appearance corruptions. In this paper, we propose a probabilistic index histogram to improve the discriminative power of the histogram representation. With this modeling, an input frame is translated into an index map whose entries indicate indexes to a separate bin. Based on the index map, we introduce spatial information and the bin-ratio dissimilarity in histogram comparison. The proposed probabilistic indexing technique, together with the two robust measurements, greatly increases the discriminative power of the histogram representation. Both qualitative and quantitative evaluations show the robustness of the proposed approach against partial occlusion, noisy and clutter background.

1 Introduction

Appearance model is one of the most important issues in object tracking. Generally speaking, the appearance model can mainly be divided into two types: histogram [1] and non-parametric description [2],[3],[4],[5], [6],[7]. Histogram-based models, which naturally capture the global statistic information of the target region, are one of the most popular models. This is due to their robustness against target deformation and noises.

However, because the color histogram is a statistic description of the target region, it loses the spatial information and not robust to background disturbance and occlusion. Also the traditional ways obtain the histogram bin by equally dividing the color space. However this division can neither accurately nor efficiently encode the color distribution of the target region.

The above drawbacks of histogram representation limit its application in visual tracking. In order to address the above issues, we propose a probabilistic index histogram with spatial distance and cross bin-ratio dissimilarity measurement. The main contributions of the proposed algorithm are summarized as follows:

1. We propose a probabilistic index histogram as the appearance model. Instead of obtaining the histogram bins by equally dividing the color space, we define each bin adaptively as a palette. Using the palette indexing theory [8], each bin is considered as a color probabilistic distribution. An image is then translated into an index map whose entries are the bin number the pixels fall in.
2. For the probabilistic index histogram, we propose an efficient *spatial distance* between two bins. The spatial distance improves the matching accuracy by capturing spatial layout for the histogram.
3. Instead of using traditional distances (e.g., Bhattacharyya distance) for comparing two histograms, we use the cross *bin-ratio dissimilarity*, which is previously proposed for category and scene classification [9], to improve the robustness to background clutters and partial occlusion.

2 Related work

Vast works have been done to increase the robustness of the histogram representation. In [10], oriented kernels are adopted to take the spatial information into consideration. Birchfiled et al. [11] introduce the spatial mean and covariance of the pixel positions of the given bin. In [12], Earth Mover's Distance (EMD) which is robust to illumination changes is employed to weighted the similarity of two histograms. In [13], object is represented by multiple image fragments and the histograms are compared with the corresponding image patch histogram.

Our method is different from the above histogram representations in both representation and similarity measurement. We model each bin as a palette and propose a new probabilistic histogram. Our probabilistic histogram code the color distribution more accurately than equally dividing the color space as histogram bin. With this modeling, an input frame is translated into an index map. Based on the index map, we introduce a spatial distance to compare spatial layout of the pixels falling in the same histogram bin. We also introduce the cross bin-ratio dissimilarity to compute the similarity of two histograms. This measurement together with the spatial distance enhances the robustness of histogram representation against occlusion, noisy and background cluttering.

The rest of this paper is structured as follows. In Section 3, the proposed algorithm is detailed. Experimental results are presented in Section 4, and Section 5 is devoted to conclusions.

3 Index histogram

3.1 Palette indexing

An efficient way to represent an image is to define it as an index matrix. Each entry in the matrix indicates index to a separate palette. The palettes are possible colors of the target region. By the definition of index matrix, image pixels corresponding to a palette share the same index. The image structure is better

captured by carefully analyzing the index map. Let I be a $M \times N$ image and $\{L_s\}_{s=1}^{m}$ be the palettes. The index for each pixel $x_{i,j}$ is represented as $d_{i,j}$, where i, j is the location of the pixel. The palette L is a table of m color or feature. For the color image, $\{L_s\} = \mu_s$ can be an $[RGB]$ vector. The index $d_{i,j}$ of each pixel points to the palette the pixel belongs to.

Instead of including all the image color in the palette, each palette L_s is defined as a Gaussian distribution and the probability of a pixel $x_{i,j}$ belonging to a certain palette L_s is formulated by a Gaussian distribution:

$$p(x_{i,j}|L_s) = \phi(x_{i,j} : \mu_s, \Sigma_s) \tag{1}$$

where μ_s, Σ_s are the mean and covariance of Gaussian distribution $\phi(.)$. Through maximizing the probability each pixel belongs to all the palettes, each entry in the index map can be obtained.

3.2 Probabilistic indexing histogram

Following the idea of palette indexing, we model each histogram bin as a color palette. Let B_s be the s^{th} histogram bin, $d_{i,j}$ indicate the bin the pixel $x_{i,j}$ falls in. Given an image I, the learning process aims to obtain the $d_{i,j}$ and B_d simultaneously. These two parameters can be obtained through maximizing the posterior probability $p(x|d, B)$. After treating the index variable d as hidden variables and bin B as the model parameters, $p(x|d, B)$ can be expressed as: $p(x|B) = \sum_d \ p(x, d|B)$. Unfortunately, this optimization is intractable, an approximate method is needed. The most popular approximation method is the variational method [14]. In the method, an alternative cost, *free energy* $\mathbb{E}$, is defined instead of directly maximizing $p(x|d, B)$:

$$\mathbb{E} = \sum_d \ q(d) \log \frac{q(d)}{p(x, d|B)} = \sum_d \ q(d) \log q(d) - \sum_d \ q(d) \log p(x, d|B) \tag{2}$$

where $q(d)$ can be an arbitrary distribution. If we define $q(d)$ as $p(x|B, d)$, $\mathbb{E}$ equals to $-log\ p(x|B)$. Using the Jensen's inequality, it can be shown that $\mathbb{E} \geq -log\ p(x|B)$. So the lower bound of $\mathbb{E}$ is the posterior probability $p(x|d, B)$ that we need to optimize. Using the variational method in [14], the free energy can be efficiently optimized using an iterative algorithm.

In order to minimize the free energy $\mathbb{E}$, we fix p and optimize q under the constraint $\sum_{i,j} q(d_{i,j}) = 1$. After minimizing the free energy $\mathbb{E}$, q is obtained as

$$q(d_{i,j}) \propto \ p(d_{i,j}) p(x_{i,j}|d_{i,j}, B) \tag{3}$$

where $p(d_{i,j})$ is the prior distribution, and $p(x_{i,j}|d_{i,j}, B)$ is defined in Equ.(1). Then the bin parameters $B = \{\mu_s, \Sigma_s\}_{s=1}^{m}$ are estimated by minimizing the free energy $\mathbb{E}$ while keeping $q(d_{i,j})$ fixed:

$$\mu_s = \frac{\sum_{i,j} q(d_{i,j} = s) x_{i,j}}{\sum_{i,j} q(d_{i,j} = s)} \tag{4}$$

$$\Sigma_s = \frac{\sum_{i,j} q(d_{i,j} = s)[x_{i,j} - \mu_s][x_{i,j} - \mu_s]^T}{\sum_{i,j} \ q(d_{i,j} = s)} \tag{5}$$

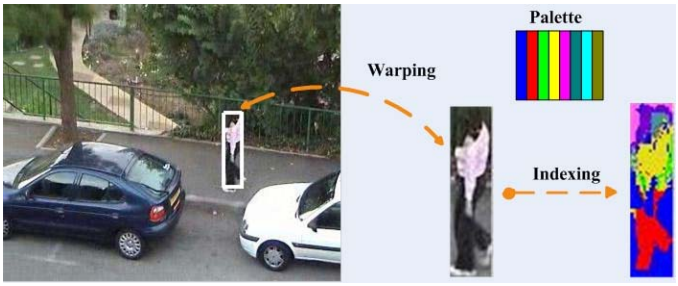


Fig. 1. Probabilistic index histogram.

These two steps are conducted iteratively until convergence. The results are probabilistic histogram whose bins are modeled as $B = \{\mu_s, \Sigma_s\}_{s=1}^{m}$. The index of the pixel $x_{i,j}$ can be obtained through minimizing the Mahalanobis distance between each pixel $x_{i,j}$ and the histogram bin:

$$d_{i,j} = \arg\min_{s}((x_{i,j} - \mu_s)\Sigma_s^{-1}(x_{i,j} - \mu_s)) \tag{6}$$

Fig.1 illustrates the result of probabilistic histogram and index map. Each color in the palette represents the mean of each histogram bin. For the clarity of illustration, each histogram bin is assigned a distinctive color instead of the original mean. Different color in the target region corresponds to different bins the pixels belong to. From Fig.1, the image can accurately be coded with the probabilistic index histogram.

3.3 Spatial distance

The histogram representation provides rich statistic information at the cost of losing spatial layout of pixels falling into the same histogram bin. However, the index map of the target region obtained using Equ.(5)(6) captures the image structure. Specifically, the distribution of pixel position of the same index is an efficient way to represent the spatial layout of the histogram bin. Motivated by this observation, we model the spatial layout of the s^{th} histogram bin using the spatial mean $\mu_{a,s}^T$ and covariance $\Sigma_{a,s}^T$ of the pixel position $a_{i,j}$ of index s,

$$\begin{aligned} \mu_{a,s}^T &= \frac{\sum_{i,j}\ a_{i,j}\delta(d_{i,j} - s)}{\sum_{i,j}\ \delta(d_{i,j} - s)} \\ \Sigma_{a,s}^T &= \frac{\sum_{i,j}\ [a_{i,j} - \mu_{a,s}^T]^T[a_{i,j} - \mu_{a,s}^T]\delta(d_{i,j} - s)}{\sum_{i,j}\ \delta(d_{i,j} - s)} \end{aligned} \tag{7}$$

where δ is the Kronecker function such that $\delta(d_{i,j} - s) = 1$ if $d_{i,j} = s$ and $\delta(d_{i,j} - s) = 0$ otherwise.

The weight of each histogram bin contributes to the whole spatial distance is proportional to the number of pixels in the index:

$$\omega_s = \frac{\sum_{i,j}\ \delta(d_{i,j} - s)}{\sum_s \sum_{i,j}\ \delta(d_{i,j} - s)} \tag{8}$$

Given a candidate region, the spatial mean $\mu^C_{a,s}$ and covariance $\Sigma^C_{a,s}$ of the s^{th} bin can be computed accordingly. The spatial distance between the target histograms H^T and a candidate histograms H^C is formulated as follows:

$$SD(H^T, H^C) = \sum_s \omega_s \exp\{-\frac{1}{2}[\mu^C_{a,s} - \mu^T_{a,s}]^T((\Sigma^T_{a,s})^{-1} + (\Sigma^C_{a,s})^{-1})[\mu^C_{a,s} - \mu^T_{a,s}]\} \tag{9}$$

3.4 Cross bin-ratio dissimilarity

A widely used method to compare the target histogram H^T and candidate histogram H^C is the Bhattacharyya distance (e.g., in [1]):

$$\rho(H^T, H^C) = \sum_{u=1}^{m} \sqrt{h^T(u)h^C(u)} \tag{10}$$

However, this measurement only considers bin to bin information and loses the cross bin interaction. In addition, as the target region is usually represented with a rectangle, it is often corrupted by the background clutters and occlusion part that are irrelevant to the target. As shown in Fig.1, the histogram bin represented with blue is obviously the background and the pixels falling into this bin account for a large portion of the target region. Such background information and occluded part will introduce noises into histogram representation and which in turn brings the inaccurate matching. In order to overcome these drawbacks, we introduce a cross bin-ratio dissimilarity measurement.

Let h be an m-bin histogram. A ratio matrix W is defined to capture the cross bin relationship. Each element in the matrix is (h_u/h_v) which measure the relation between bin $h(u)$ and $h(v)$. The whole ratio matrix is written as follows:

$$W = \left(\frac{h_u}{h_v}\right)_{u,v} = \begin{bmatrix} \frac{h_1}{h_1} & \frac{h_2}{h_1} & \cdots & \frac{h_m}{h_1} \\ \frac{h_1}{h_2} & \frac{h_2}{h_2} & \cdots & \frac{h_m}{h_2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{h_1}{h_m} & \frac{h_2}{h_m} & \cdots & \frac{h_m}{h_m} \end{bmatrix} \tag{11}$$

With the definition of the ratio matrix, we compare the vth bin between two histogram H^T and H^C using dissimilarity M_v. M_v is defined as the sum of squared difference between the vth rows of corresponding ratio matrix:

$$M_v(H^T, H^C) = \sum_{u=1}^{m} (\frac{h^T_u}{h^T_v} - \frac{h^C_u}{h^C_v})/(\frac{1}{h^T_v} + \frac{1}{h^C_v}) \tag{12}$$

where $\frac{1}{h^T_v} + \frac{1}{h^C_v}$ is normalization term to avoid the instability problem when h^T_v and h^C_v close to zero. From the above definition, the influence of the clutter or occlusion part bin is weakened by the ratio operation. Thus this measurement is robust to background clutter and occlusion.

We simplify M_v using the $\mathbb{L}_2$ normalization $\sum_{k=1}^{m} h^2(k) = 1$ and formulate the cross bin-ratio dissimilarity M between histogram H^T and H^C as follows:

$$M(H^T, H^C) = \sum_{v=1}^{m} M_v(H^T, H^C) = \sum_{v=1}^{m} (1 - \frac{h_v^T h_v^C}{(h_v^T + h_v^C)^2} \|H^T + H^C\|_2^2) \quad (13)$$

3.5 Bayesian state inference for object tracking

In this paper, the object is localized with a rectangular window and its state is represented using a six dimension affine parameter $X_t = (t_x, t_y, \theta, s, \alpha, \beta)$ where (t_x, t_y) denote the 2-D translation parameters and $(\theta, s, \alpha, \beta)$ are deforming parameters. Given the observation I_t, the goal of the tracking is to infer X_t. This inference can be cast as a Bayesian posterior probability inference process [15],

$$p(X_t|I_t) \propto p(I_t|X_t) \int p(X_t|X_{t-1}) p(X_{t-1}|I_{t-1}) dX_{t-1} \quad (14)$$

where $p(I_t|X_t)$ is the observation model and $p(X_t|X_{t-1})$ represents the dynamic model. A particle filter [15] is used to approximate the posterior probability with a set of weighted samples. We use a Gaussian distribution to model the state transition distribution. The observation model $p(I_t|X_t)$ reflects the similarity between the candidate histogram of state X_t and target histogram:

$$p(I_t|X_t) = \exp\{-\frac{1}{2\sigma^2}(1 - SD(H^T, H^C)\} * \exp\{-\frac{1}{2\sigma^2} * \alpha M(H^T, H^C)\} \quad (15)$$

where σ is the observation variance and α is a weighting factor to balance the influence of spatial distance and cross bin-ratio dissimilarity. If we draw particles from the state transition distribution, the weight $\mathbb{W}_t$ of each sample X_t can be evaluated by the observation likelihood $p(I_t|X_t)$. Then we use a *maximum a posterior* (MAP) estimate to obtain the state of the object at each frame.

4 Experiments

In order to validate the effectiveness of our proposed method, we perform a number of experiments on various sequences. Comparisons with other algorithms are also presented to further show the superiority of our approach. To give a fair comparison with the mean shift algorithm which can only deal with scale changes, we only consider the scale change of the affine parameter. The parameters are set to $\Sigma = \mathrm{diag}(5^2, 5^2, 0.01^2, 0, 0.001^2, 0)$, and 420 particles are used.

Experiment 1. The first experiment is conducted to test the influence of the spatial distance and cross bin-ratio dissimilarity respectively. Also comparison with the color histogram based mean shift algorithm [1] is presented. The sequence is a woman partially occluded by cars. The cars and some background are similar in appearance to the woman. Fig.2 (a) shows the tracking results of mean shift. Obviously the mean shift algorithm is distracted by the cars and

(a) Mean shift

(b) The spatial distance only

(c)The cross bin-ratio dissimilarity only

(d) The proposed algorithm

Fig. 2. The tracking results of Experiment 1.

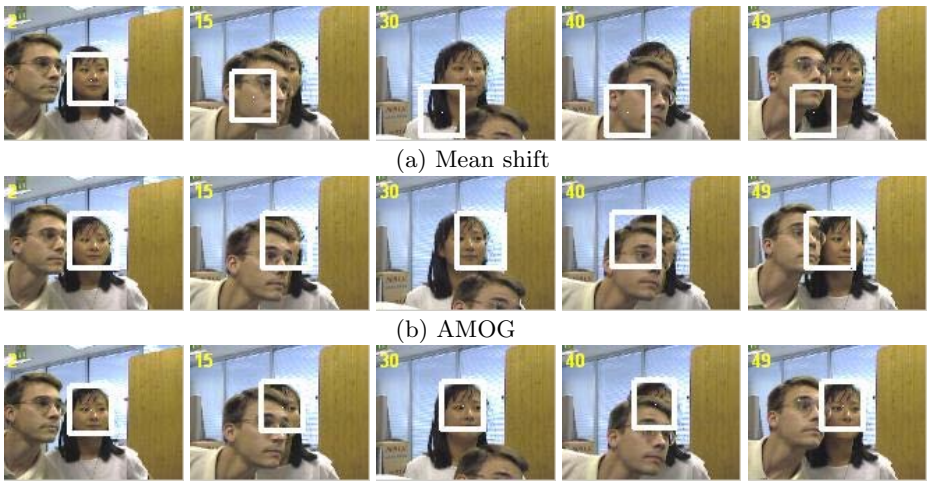
similar background and can not deal with partial occlusion well. The results in Fig.2 (b) and Fig.2 (c) illustrate that both spatial distance and cross bin-ratio dissimilarity improve the tracking results. However, only one term can not always provide satisfying results. From the tracking results in Fig.2 (d), our proposed algorithm which combines the spatial distance and cross bin-ratio dissimilarity successfully tracks all the frames and provides accurate results.

Tracking approach	Mean shift	Spatial only	Cross bin-ratio only	Our approach
RMSE of Position	14.7018	7.9759	6.6581	3.2451

Table 1. Quantitative results for Experiment 1

A quantitative evaluation of four algorithms is presented in Table.1. We compute the RMSE (root mean square error) between the estimated position and the groundtruth. Here the groundtruth is marked by hand. The results in Table.1 validate that our algorithm with spatial distance and cross bin-ratio dissimilarity achieve the most accurate results.

Experiment 2. In the second experiment, we test the performance of our algorithm in handling partial occlusion and background distraction. We compare our algorithm with other two algorithms. One is the color histogram based mean shift [1] algorithm which only consider the statistical information of object. The other one is a popular parametric description algorithm [5], which adopts an *adaptive mixture of Gaussians* (AMOG) as the appearance model. From Fig.3(a), the mean shift algorithm is distracted away by another face with similar color and



(a) Mean shift

(b) AMOG

(c) Our proposed algorithm

Fig. 3. The tracking results of the Experiment 2.

can not recover anymore. The AMOG also can not provide good results. On the contrary, our algorithm is capable of tracking the object through all the frames even though the face endures severely occlusion and background distraction.

Experiment 3. The third experiment aims to test the robustness of our algorithm against clutter background and scene blurring. As shown from Fig.4(a), the mean shift algorithm quickly drifts away and can not capture the object any more. This is mainly because the nearby background and the object share the similar color histogram statistics. The tracking results in Fig.4(b) show that the AMOG also can not tackle the clutter background. However the good tracking results in Fig.4(c) illustrate that our algorithm is robust against clutter background and scene blurring.

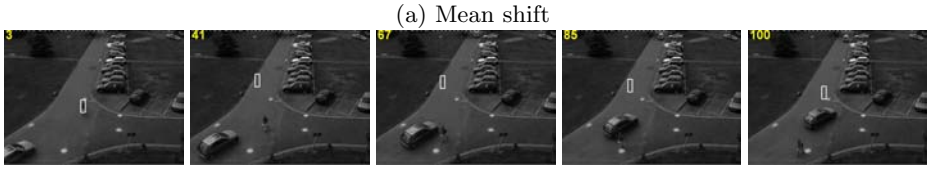
Experiment 4. In the last experiment, we test our algorithm on a more challenging sequence. In this sequence, a car moves in a noisy background. The nearby background is so noisy that the car can not easily be located even by eyes. Fig.5 presents the tracking results of three algorithms. As shown in Fig.5(c), the noisy background poses no challenges for our algorithms. However both mean shift and AMOG encounter troubles in the extremely noisy background.

Tracking approach	Mean shift		AMOG		Our approach	
Evaluation method	RMSE	STF	RMSE	STF	RMSE	STF
Second sequence	20.8207	14/56	6.4717	43/56	2.5356	56/56
Third sequence	85.8135	1/100	92.2919	4/100	5.5866	100/100
Fourth sequence	15.6012	3/183	18.5673	27/183	3.7865	183/183

Table 2. Quantitative results for last three sequences

A quantitative evaluations of the last three sequences are also given in Table 2 to further demonstrate the superiority of our algorithm. The evaluation is

(a) Mean shift



(b) AMOG

(c) Our proposed algorithm

Fig. 4. The tracking results of Experiment 3.

comprised of the following two aspects: RMSE, STF(the number of successfully tracked frames and the tracking is defined as failure if the center of the window is not in the object). From the results in Table 2, we make the following conclusions: (1) The mean shift and the AMOG algorithm are only suitable for the tracking when the appearance of the object is different from the background. Both these two algorithms can not deal with occlusion, noisy and clutter background well; (2) The spatial distance and the cross bin-ratio dissimilarity based on the probabilistic index histogram make our approach robust to occlusion, noisy and clutter background. As a result, our proposed approach is an effective way to improve the discriminative power of the traditional histogram representation.

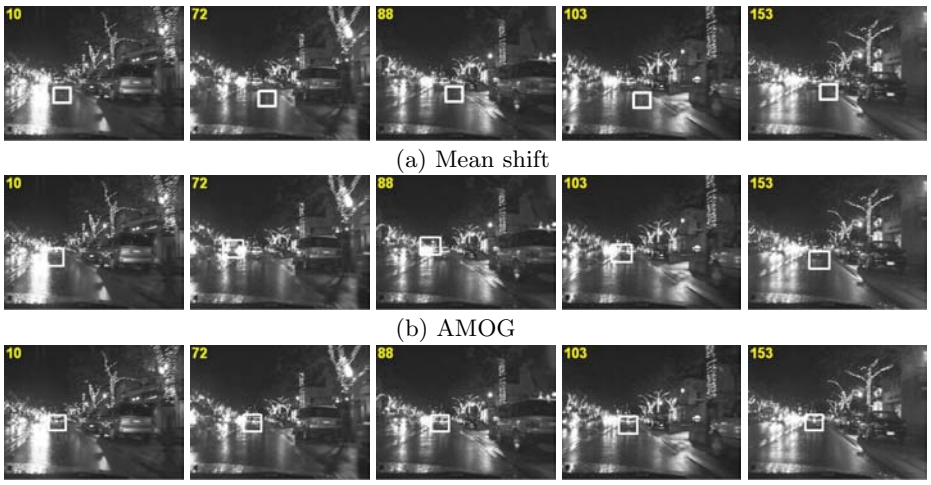
5 Conclusions

In this paper, we propose a probabilistic index histogram to increase the robustness of the color histogram representation. Our new histogram representation, together with spatial distance and cross bin-ratio dissimilarity, greatly increase the discriminative power of the histogram representation. In experiments on several challenging sequences validate the claimed contributions

Acknowledgement. This work is partly supported by NSFC (Grant No. 60825204 and 60935002) and the National 863 High-Tech R&D Program of China (Grant No. 2009AA01Z318).

References

1. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking (2003)
2. Black, M., Jepson, A.: Eigentracking: Robust matching and tracking of articulated objects using view-based representation. In Proc. ICCV (1995) 329–342



(a) Mean shift

(b) AMOG

(c) Our proposed algorithm

Fig. 5. The tracking results of Experiment 4.

3. Jepson, A., Fleet, D., El-Maraghi, T.: Robust online appearance models for visual tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 1296C1311
4. Black, M., Fleet, D., Yacoob, Y.: A framework for modeling appearance change in image sequence. In Proc. ICCV (1998) 660–667
5. Zhou, S., Chellappa, R., Moghaddam, B.: Visual tracking and recongnition using appearance-adaptive models in particles filters. IEEE Transaction on Image Processing **13** (2004) 1491C1506
6. Lim, J., Ross, D., Lin, R., Yang, M.: Incremental learning for visual tracking. NIPS (2005) 793–800
7. Perez, P., Hue, C., Vermaak, J., Gangnet, M.: Color-based probabilistic tracking (2002)
8. Jojic, N., Caspi, Y.: Capturing image structure with probabilistic index maps (2004)
9. Xie, N., Ling, H., Hu, W., Zhang, Z.: Use bin-ratio information for category and scene classification (2010)
10. Georgescu, B., Meer, P.: Point matching under large image deformations and illumination changes. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 674C689
11. Georgescu, B., Meer, P.: Spatiograms vs. histograms for region based tracking. IEEE Conf. on Computer Vision and Pattern Recognition (2005)
12. Zhao, Q., Brennan, S., Tao, H.: Differential emd tracking. In Proc. ICCV (2007)
13. Adam, A., Rivlin, E., Shimshoni, I.: Robust fragmentsbased tracking using the integral histogram. In IEEE Conf. Computer Vision and Pattern Recognition (2006)
14. Jordan, M., Ghahramani, Z., Jaakkola, T., Saul, L.: An introduction to variational methods for graphical models. in Learning in Graphical Models, M. I. Jordan, Ed. Kluwer Academic Publishers (1998)
15. Isard, M., Blake, A.: Contour tracking by stochastic propagation of conditional density. In Proc. ECCV **2** (1996) 343–356