Effective Appearance Model and Similarity Measure for Particle Filtering and Visual Tracking

Hanzi Wang, David Suter, and Konrad Schindler

Institute for Vision Systems Engineering,
Department of Electrical and Computer Systems Engineering,
Monash University, Clayton Vic. 3800, Australia
{hanzi.wang, d.suter, konrad.schindler}@eng.monash.edu.au

Abstract. In this paper, we adaptively model the appearance of objects based on Mixture of Gaussians in a joint spatial-color space (the approach is called SMOG). We propose a new SMOG-based similarity measure. SMOG captures richer information than the general color histogram because it incorporates spatial layout in addition to color. This appearance model and the similarity measure are used in a framework of Bayesian probability for tracking natural objects. In the second part of the paper, we propose an Integral Gaussian Mixture (IGM) technique, as a fast way to extract the parameters of SMOG for target candidate. With IGM, the parameters of SMOG can be computed efficiently by using only simple arithmetic operations (addition, subtraction, division) and thus the computation is reduced to linear complexity. Experiments show that our method can successfully track objects despite changes in foreground appearance, clutter, occlusion, etc.; and that it outperforms several color-histogram based methods.

1 Introduction

Visual tracking in unconstrained environments is one of the most challenging tasks in computer vision because it has to overcome many difficulties arising from sensor noise, clutter, occlusions and changes in lighting, background and foreground appearance etc. Yet tracking objects is an important task with many practical applications such as smart rooms, human-computer interaction, video surveillance, and gesture recognition. Generally speaking, methods for visual tracking can be roughly classified into two major groups: deterministic methods and stochastic methods.

In deterministic methods (for example, the Mean Shift (MS) tracker [1]), the target object is located by maximizing the similarity between a template image and the current image. The localization is implemented by iterative search. These methods are computationally efficient, but they are sensitive to background distraction, clutter, occlusion, etc. Once they lose the target object, they can not recover from the failure on their own. This problem can be mitigated by stochastic methods, which maintain multiple hypotheses in the state space and in this way, achieve more robustness. For example, the Particle Filter (PF) [2, 3, 4] has been widely applied in visual tracking in recent years.

A particle filter tracks multiple hypotheses simultaneously and weights them according to a similarity measure (i.e., the observation likelihood function). This paper

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part III, LNCS 3953, pp. 606 – 618, 2006.

is essentially concerned with devising and calculating this likelihood function/similarity measure. Visual similarity can be measured using many features such as intensity, color, gradient, contour, texture, or spatial layout. A popular feature is color [1, 2, 4, 5, 6], due to its simplicity and robustness (against scaling, rotation, partial occlusion, and non-rigid deformation). Usually, the appearance of a region is represented by its color histogram, and the distance between the normalized color histograms of two regions is measured by the Bhattacharyya distance [2, 4].

Despite its popularity, the color histogram also has several disadvantages:

1) The spatial layout information of a tracked object is completely ignored (see figure 1(a)). As a result, a tracker based on color histograms is easily confused when two objects with similar colors but different spatial distributions get close to each other. An ad-hoc solution is to manually split the tracked region into several sub-regions (e.g., [4, 7]).
2) Since the appearance of the target object is reduced to a global histogram, the similarity measure (e.g., the Bhattacharyya coefficient) is not discriminative enough (see Fig. 1) [8].
3) For a classical color histogram based particle filter, the construction of the histograms is a bottleneck. The computation is quadratic in the number of samples.

In order to overcome the disadvantages of color histograms, we describe a Spatial-color Mixture of Gaussians (called SMOG) appearance model and propose a SMOG-based similarity measure in Sect. 2. The main advantage of SMOG over color histograms and general Gaussian Mixtures is in that both the color information and the spatial layout information are utilized in the objective function of SMOG. Therefore, the SMOG-based similarity measure is more discriminative.

When SMOG and the SMOG-based similarity measure are used in particle filters, one major bottleneck is the extraction of the parameters (weight, mean, and covariance) of SMOG for each particle. In Sect. 3, we propose an Integral Gaussian Mixture (IGM) technique as a fast way to extract these parameters and which also requires less memory storage than the integral histogram [9].

In Sect. 4, experiments showing the advantages of our method over other popular methods are provided. We summarize the paper in Sect. 5.

2 SMOG for Particle Filters

2.1 A Brief Review of the Particle Filter

Denoting by X_t and Y_t the hidden state and the observation respectively at time t. The goal is to estimate the posterior probability density function (pdf) $p(X_t)$ of the target object state given all available observations up to time t: $Y_{1:t}=\{Y_i, i=1, \ldots,t\}$. Employing the first-order Markovian assumption $p(X_t|X_{1:t-1}) = p(X_t|X_{t-1})$, the posterior distribution of the state variable can be formulated as follows:

$$p(X_t|Y_{1:t}) \propto L(Y_t \mid X_t) \int p(X_t|X_{t-1}) p(X_{t-1}|Y_{1:t-1}) dX_{t-1} \quad (1)$$

Given the dynamic model $p(X_t|X_{t-1})$ and the observation likelihood model $L(Y_t|X_t)$, the posterior pdf distribution in Eq (1) can be recursively calculated.

The particle filter approximates the posteriori distribution $p(X_t|Y_{1:t})$ based on a finite set of random particles and associated weights $\{X_t^{(j)}, W_t^{(j)}\}_{j=1}^{M}$. If we draw particles from an importance density, i.e., $X_t^{(j)} \sim q(X_t^{(j)}|X_{t-1}^{(j)}, Y_{1:t})$, the weights of new particles become:

$$W_t^{(j)} \propto \frac{L(Y_t|X_t^{(j)})p(X_t^{(j)}|X_{t-1}^{(j)})}{q(X_t^{(j)}|X_{t-1}^{(j)}, Y_{1:t})} \tag{2}$$

Then, the state estimate of the object at each frame can be obtained by either the mean state or a maximum a posteriori (MAP) estimate [10].

The observation likelihood function $L(Y_t|X_t)$ plays an important role in the particle filter. It determines the weights of particles and thereby could significantly influence the performance [11]. The likelihood function mainly affects the particle filter by the following ways:

1) It affects the way particles are re-sampled. Re-sampling is necessary to decrease the number of low weighted particles and to increase the ones with more potential particles. Particles are re-sampled according to their weights.
2) It affects the state estimate $\hat{X}_t$ of the target object.

Two popular likelihood function categories are: contour-based models (e.g., [12]) and color-based models (e.g. [1, 2, 4, 6]). Although the contour-based model can accurately describe the shape of a target, it performs poorly in clutter and the time complexity is high. In the color-based model, a color histogram (due to its robustness to noise, rotation, and partial occlusion, etc.) is frequently employed with the Bhattacharyya coefficient as a similarity measure. However, color histogram has some limitations, as we show next.

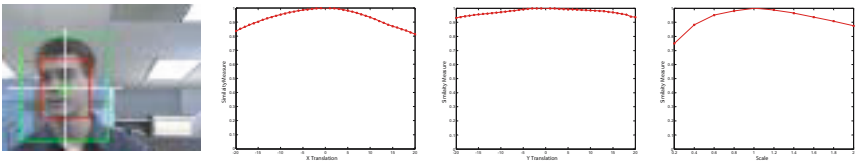
2.2 Limitations of Color-Histogram Based Similarity Measure

We illustrate the main disadvantage of the color histogram based similarity measure: it lacks information about the spatial layout of the target object, and is thus not discriminative enough.

Denote by $\phi_{O_t} = \{\phi_{O_t}^{(u)}\}_{u=1,\ldots,m}$ and $\phi_{O_v} = \{\phi_{O_v}^{(u)}\}_{u=1,\ldots,m}$ respectively the m-bin normalized color histograms of target model o_t and the target candidate o_v, the Bhattacharyya coefficient (i.e., the similarity measure) between the reference region and candidate region is:

$$\rho\left(\phi_{O_t}, \phi_{O_v}\right) = \sum_{u=1}^{m} \sqrt{\phi_{O_t}^{(u)} \phi_{O_v}^{(u)}} \tag{3}$$

In Fig. 1, we track a face comprising pixels within a red rectangle region in a video sequence from http://vision.stanford.edu/~birch/headtracker/seq/. Target candidates



(a) (b) (c) (d)

Fig. 1. Color-histogram based similarity measure. The score of the similarity measure over (b) x-translation; (c) y-translation; and (d) scaling. (see text below and compare with Fig. 2).

are generated by translating the rectangle from -20 to 20 horizontally or vertically, and by scaling the rectangle by a factor of 0.2 (the smaller green rectangle inside the target model) to 2 (the larger green rectangle inside the target model) in steps of 0.2. We use 8x8x8 color histogram bins. From Fig. 1, we can see that the similarity measure by Eq. (3) obtains very similar scores for different target candidates, and does not discriminate well between different candidate regions.

2.3 SMOG: A Joint Spatial-Color Appearance Model

Both the appearance model and the similarity measure are very important to the performance of particle filters. The color histogram, as described above, is one popular appearance model. Other popular models for foreground and/or background appearance include: the Gaussian [13], the kernel density [14, 15] and the MOG (Mixture of Gaussians) based appearance model [10, 16, 17, 18, 19, 20]. For example, [13] represented humans by blobs and modeled each blob by a Gaussian model.

The kernel density based model is robust to noise and does not require the calculation of parameters (such as weights, mean and covariance of the Gaussian model) but it is computationally expensive and requires a large storage space. It is also not trivial to update the appearance changes. The disadvantage of the general MOG-based model is that it treats each pixel independently without using any spatial information. Moreover, it requires setting the number of Gaussians and a learning rate. Despite these limits, it is popular because (1) it can model the multi-modal distribution of the appearance; (2) it is computationally efficient; (3) it is easy to adapt to the changes of the appearance; and (4) it does not require a large storage space.

We model the appearance of an object with a joint spatial-color mixture of Gaussians. We refer to this approach as SMOG. We denote by $S_i=(x_i, y_i)$ and $C_i=\{C_i^j\}_{j=1,\ldots,d}$ respectively the spatial feature (i.e., the 2D coordinates) and the color feature with d color channels (in RGB color space, $C_i=\{R_i,G_i,B_i\}$ and d=3) at pixel x_i. Thus, we can write the features of x_i as the Cartesian product of its position and color: $x_i = (S_i,C_i)$. We assume that the spatial feature (S) and the color feature (C) are independent to each other. For the mean and the covariance of the lth mode of the

Gaussian Mixtures, we have $\mu_{t,l} = (\mu_{t,l}^{\mathrm{S}}, \mu_{t,l}^{\mathrm{C}})$ and $\Sigma_{t,l} = (\Sigma_{t,l}^{\mathrm{S}}, \Sigma_{t,l}^{\mathrm{C}})$. The estimated density at the point x_i in the joint spatial-color space can be written as:

$$p_o(x_i) = \sum_{l=1}^{k} \omega_{t,l} \frac{\exp\left\{-\frac{1}{2}(\mathrm{S}_i - \mu_{t,l}^{\mathrm{S}})^T (\Sigma_{t,l}^{\mathrm{S}})^{-1} (\mathrm{S}_i - \mu_{t,l}^{\mathrm{S}})\right\}}{2\pi |\Sigma_{t,l}^{\mathrm{S}}|^{1/2}} \frac{\exp\left\{-\frac{1}{2}(\mathrm{C}_i - \mu_{t,l}^{\mathrm{C}})^T (\Sigma_{t,l}^{\mathrm{C}})^{-1} (\mathrm{C}_i - \mu_{t,l}^{\mathrm{C}})\right\}}{(2\pi)^{d/2} |\Sigma_{t,l}^{\mathrm{C}}|^{1/2}} \tag{4}$$

2.4 SMOG-Based Similarity Measure

We model the appearance of a target object O_t by SMOG with k modes. We initialize the parameters of SMOG for a target object $\{\omega_{t=1,l}^{O_t}, \mu_{t=1,l}^{\mathrm{S},O_t}, \mu_{t=1,l}^{\mathrm{C},O_t}, \Sigma_{t=1,l}^{\mathrm{S},O_t}, \Sigma_{t=1,l}^{\mathrm{C},O_t}\}_{l=1,\ldots,k}$ by a K-means algorithm followed by a standard EM algorithm. Once we obtain the parameters of the target object, we either update these parameters in an "exponential forgetting" way or keep the parameters (if we detect that it is occluded by other objects) in the following frames (t=2,3…). At time t, we sample M particles (i.e., target candidates O_v) and evaluate the likelihood function in Eq. (1) for each particle. The parameters of each target candidate $\{\omega_{t,l}^{O_v}, \mu_{t,l}^{\mathrm{S},O_v}, \mu_{t,l}^{\mathrm{C},O_v}, \Sigma_{t,l}^{\mathrm{S},O_v}, \Sigma_{t,l}^{\mathrm{C},O_v}\}_{l=1,\ldots,k}$ are calculated by:

1. Calculate the Mahalanobis distances between pixels $\{x_i\}$ in the target candidate O_v=$\{x_i\}_{i=1,\ldots N}$ to each mode of SMOG of the target object O_t in color space:

$$D_l^2(\mathrm{C}_i, \mu_{t,l}^{\mathrm{C},O_t}, \Sigma_{t,l}^{\mathrm{C},O_t}) = (\mathrm{C}_i - \mu_{t,l}^{\mathrm{C},O_t})^T (\Sigma_{t,l}^{\mathrm{C},O_t})^{-1} (\mathrm{C}_i - \mu_{t,l}^{\mathrm{C},O_t}) \tag{5}$$

2. Label the pixels satisfying ANY($|\mathrm{D}_l|_{l=1,\ldots,\mathrm{k}} \le 2.5$) with the number of the mode to which the Mahalanobis distance is the least. For other pixels, label them with zero.

$$LB(x_i) = \arg\min_l |D_l| \tag{6}$$

3. Calculate the parameters $\{\omega_{t,l}^{O_v}, \mu_{t,l}^{\mathrm{S},O_v}, \mu_{t,l}^{\mathrm{C},O_v}, \Sigma_{t,l}^{\mathrm{S},O_v}, \Sigma_{t,l}^{\mathrm{C},O_v}\}_{l=1,\ldots,k}$ of the target candidate by:

$$\begin{aligned}
\omega_{t,l}^{O_v} &= \left(\sum_{i=1}^{N} \delta(LB(x_i) - l)\right) \Big/ \left(\sum_{l=1}^{k}\sum_{i=1}^{N} \delta(LB(x_i) - l)\right) \\
\mu_{t,l}^{O_v} &= (\mu_{t,l}^{\mathrm{S},O_v}, \mu_{t,l}^{\mathrm{C},O_v}) = \left(\sum_{i=1}^{N} x_i \delta(LB(x_i) - l)\right) \Big/ \left(\sum_{i=1}^{N} \delta(LB(x_i) - l)\right) \\
\Sigma_{t,l}^{O_v} &= (\Sigma_{t,l}^{\mathrm{S},O_v}, \Sigma_{t,l}^{\mathrm{C},O_v}) = \left(\sum_{i=1}^{N} \left(x_i - \mu_{t,l}^{O_v}\right)^T \left(x_i - \mu_{t,l}^{O_v}\right) \delta(LB(x_i) - l)\right) \Big/ \left(\sum_{i=1}^{N} \delta(LB(x_i) - l)\right)
\end{aligned} \tag{7}$$

where δ is the Kronecker delta function. The covariance matrix is taken to be a diagonal matrix for simplicity. One should normalize the coordinate space first so that the coordinates of pixels in the target candidate (and target object) are within the range [0, 1].

Let $\Lambda_{t,l}^{\mathrm{S}}$ and $\Lambda_{t,l}^{\mathrm{C}}$ be respectively the spatial and the color similarity measure between the lth mode of the target candidate O_v and the lth mode of the target object O_t. The SMOG-based similarity measure (as compared to the color-histogram based similarity measure in Eq. (3)) between two regions (O_v and O_t) in the joint spatial-color space is defined as:

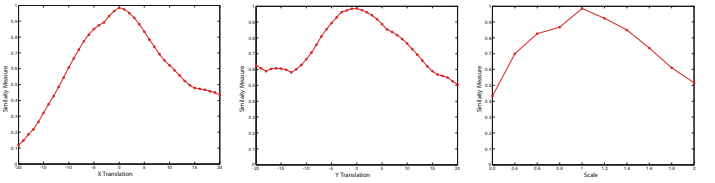
$$\Lambda(O_t, O_v) = \sum_{l=1}^{k} \Lambda_{t,l}^{\mathrm{S}} \Lambda_{t,l}^{\mathrm{C}} \tag{8}$$

where $\Lambda_{t,l}^{\mathrm{S}} = \exp\left\{-\frac{1}{2}(\mu_{t,l}^{\mathrm{S},O_v} - \mu_{t,l}^{\mathrm{S},O_t})^T (\hat{\Sigma}_{t,l}^{\mathrm{S}})^{-1} (\mu_{t,l}^{\mathrm{S},O_v} - \mu_{t,l}^{\mathrm{S},O_t})\right\}$ with $(\hat{\Sigma}_{t,l}^{\mathrm{S}})^{-1} = (\Sigma_{t,l}^{\mathrm{S},O_v})^{-1} + (\Sigma_{t,l}^{\mathrm{S},O_t})^{-1}$ and $\Lambda_{t,l}^{\mathrm{C}} = \min(\omega_{t,l}^{O_v}, \omega_{t,l}^{O_t})$.

The likelihood function in our method is given by:

$$L(\mathrm{Y}_t \mid \mathrm{X}_t) \propto \exp\left\{-\frac{1}{2\sigma_b^2}(1 - \Lambda(O_t, O_v))\right\} \tag{9}$$

where σ_{b} is the observation variance.



(a) (b) (c)

Fig. 2. The score by the SMOG-based similarity measure over (a) x-translation; (b) y-translation; and (c) scale

We repeat the experiment in Fig. 1 using SMOG. As shown in Fig. 2, the SMOG-based similarity measure (Eq. (8)) is more discriminative than the color-histogram based similarity measure in Eq. (3).

Recently, Birchfield et al. [21] proposed a method (Spatiograms), which captures the spatial information of the general histogram bins, and applied it to the Mean Shift (MS) tracker. The spatial mean and covariance *of each bin* is computed. In contrast, we consider the spatial layout and color distribution *of each mode* of SMOG. The number of the Gaussians (normally, k is set within the range from 3 to 7 in our case) is much less than the number of the histogram bins. SMOG is also more efficient in estimating density distribution of the data and in computation, and requires less storage space to build up an integral Gaussian mixtures image (as described in Sect. 3) than the integral histogram method [9].

2.5 Updating the Parameters of SMOG

We dynamically model the object appearance by updating the parameters of SMOG through a learning rate α. The assumption made here is that in the temporally neighboring frames (e.g., frame t and frame t-1), the appearance (including both spatial and color distributions) of an object does not change dramatically.

Similar to [10] and [17], we assume that the past appearance is exponentially forgotten and new information is gradually added to the appearance model.

To handle occlusion where image outliers exist, we use a heuristic way: we update the appearance only if the score of the similarity measure is larger than a threshold $\mathrm{T_u}$. When occlusion is declared (i.e., the score is less than $\mathrm{T_u}$), we stop updating the appearance model.

2.6 Choosing the Color Space

We employ the normalized color space in our method. The normalized chromaticity coordinates of (r, g, b) can be written as: r=R/(R+G+B); g=G/(R+G+B); b=R/(R+G+B). The intensity information is also exploited. Thus we use (r, g, I) as the color feature in our method.

In Fig. 3, we show an experiment illustrating the advantage of (r, g, I) over (R, G, B) color space in dealing with illumination changes. (r, g, I) color space shows more robustness to the illumination change. In contrast, the method employing (R, G, B) achieved less accurate results and lost the target at the end.

Fig. 4 shows the adaptation of the proposed method to the appearance changes by updating the appearance model in subsection 2.5. Our method succeeds in adaptation to appearance changes throughout the sequence.

Fig. 3. Tracking results employing RGB as color feature (in the first row) and *rgI* as color feature (in the second row)

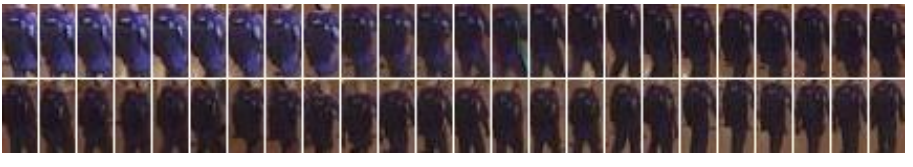


Fig. 4. The appearance of the tracked target changes with time increasing

3 Integral Gaussian Mixture for Higher Computational Efficiency

To efficiently calculate the similarity measure $\Lambda(O_t, O_v)$ (in Eq. (8)), we need to calculate $\left\{\omega_l^{O_v}, \mu_l^{S,O_v}, \Sigma_l^{S,O_v}\right\}_{l=1,\ldots,k}$ for each target candidate. One possible way, which is

usually used in the color-histogram based particle filters (such as [2, 4]), is to randomly sample a particle, and generate a target candidate, and then calculate the parameters corresponding to the candidate region. This is computationally inefficient because particles may have many overlapped regions and the same operator for each possible region can be repeated many times.

To overcome this inefficiency, integral methods exploiting rectangle features were introduced by Viola et al. [22] and more recently, were developed by Porikli [9]. In [22], a grey-level image is converted to integral image format (i.e., the value of each pixel is the sum of values of all pixels to the left and above of the current pixel). In [9], integral histogram is constructed by a recursive propagation of an aggregated histogram in a Cartesian data space.

We propose an Integral Gaussian Mixture (IGM) technique as a fast and efficient way to extract the parameters of SMOG for each particle. To calculate the parameters of the *l*th mode of a target candidate, we need to calculate $(n_l, \mu_{\mathrm{x},l}, \mu_{\mathrm{y},l}, \sigma^2_{\mathrm{x},l}, \sigma^2_{\mathrm{y},l})$, i.e., the number of pixels whose label is *l*, the spatial mean and variance values in x and y coordinates.

We can write these quantities in the following form:

$$\begin{aligned} n_l &= \sum_{i=1}^{N} \delta(LB(x_i) - l) \\ \mu_{\mathrm{x},l} &= \left(\sum_{i=1}^{N} \mathrm{x}_i \delta(LB(x_i) - l) \right) \Big/ n_l \,;\; \mu_{\mathrm{y},l} = \left(\sum_{i=1}^{N} \mathrm{y}_i \delta(LB(x_i) - l) \right) \Big/ n_l \\ \sigma^2_{\mathrm{y},l} &= \left(\sum_{i=1}^{N} \mathrm{x}_i^2 \delta(LB(x_i) - l) \right) \Big/ n_l - \mu^2_{\mathrm{x},l} ; \sigma^2_{\mathrm{y},l} = \left(\sum_{i=1}^{N} \mathrm{y}_i^2 \delta(LB(x_i) - l) \right) \Big/ n_l - \mu^2_{\mathrm{y},l} \end{aligned} \tag{10}$$

and we have

$$\omega_l = n_l \Big/ \sum_{l=1}^{k} n_l \,; \mu_l^{\mathrm{S}} = (\mu_{\mathrm{x},l},\, \mu_{\mathrm{y},l}); \Sigma_l^{\mathrm{S}} = \begin{pmatrix} \sigma^2_{\mathrm{x},l} & 0 \\ 0 & \sigma^2_{\mathrm{y},l} \end{pmatrix} \tag{11}$$

The procedure of the IGM can be described as follows:

1. Predict the region $\tilde{R}$, that includes all particles (i.e., target candidates), in the 2D image.
2. Label each pixel $x_{\tilde{i}}$ in $\tilde{R}$ by step 1 and 2 in subsection 2.4.
3. Generate a GM image whose $\tilde{i}$th pixel is given by $x_{\tilde{i}} = \{x_{\tilde{i},l}\}_{l=1,\ldots,k}$, where $x_{\tilde{i},l} = \delta(LB(x_{\tilde{i}}) - l)(1, \mathrm{x}_{\tilde{i}}, \mathrm{x}^2_{\tilde{i}}, \mathrm{y}_{\tilde{i}}, \mathrm{y}^2_{\tilde{i}})$.
4. Build an IGM image, where each pixel is the sum of values of all pixels of the GM image to the left and above of the current pixel.
5. Calculate the parameters of each target candidate by four table lookup operations, which are similar to [22].

We find that once the IGM is built, the calculation of the likelihood function is very fast. Fig. 5 gives a rough estimation of the computational time (in MATLAB code) to evaluate the likelihood function for particles. From Fig. 5, we can see that the calculation of the color histogram based similarity measure in Condensation is

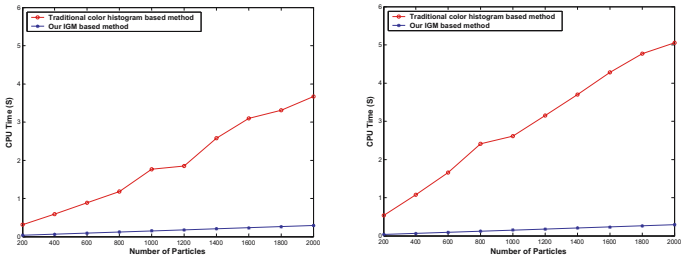


Fig. 5. The computational time v.s. the number of particles for the color histogram based method and the proposed method. Candidate region size in (b) is twice as that in (a).

computationally expensive and will be affected by both the number of particles and the size of target candidate regions. When we double the region size (Fig. 5 (b)) of the candidate region (Fig. 5 (a)), the computational time of the color histogram based Condensation increased by about 60%. In contrast, both the number of particles and the size of the target candidate regions have much less influence on the computational complexity of the proposed method: the processing time is about 10 to 20 times less than the color histogram based Condensation.

4 Experiments

We test the effectiveness of our method using a number of video sequences with different environments and conditions[1]. We compare with two popular color histogram based

Fig. 6. Tracking results of the *face* sequence with the MS tracker (first row), Condensation (second row) and our method (third row)

[1] Some demo video sequences of our method can be obtained from http://users.monash.edu.au/~hanzi

methods: the Mean Shift tracker and Condensation. Note: we employ the (r, g, I) color space for all three methods. For the Mean Shift tracker and the Condensation tracker, we use 16x16x16 color histogram bins. For both Condensation and our method, we employ a random walk dynamic model (number of particles M=200).

In Fig. 6, the human face is moving to the left and right very quickly. The illumination on the face also changes. The background scene includes clutter and material of similar color to the face. As we can see in Fig. 6, the Mean Shift tracker fails to track the face very soon; the results of Condensation are not accurate and Condensation even fails to track the face in some frames because the color histogram based similarity measure is not discriminative enough (section 2.2). In comparison, our method, which considers both color and spatial information of the target object, never loses the target and achieves the most accurate results.

Fig. 7 and Fig. 8 show situations where two humans with very similar colors get close to each other and one occludes the other. In Fig. 7, when the man's face gets close to and occludes the girl's face, the results of both the MS tracker and Condensation are greatly influenced. In Fig. 8 (a), because the color histogram based similarity ignores the spatial information, both the MS tracker and Condensation break down when two players with similar colors, but different spatial distributions, get close to each other. In contrast, our method works well in both cases. Fig. 8 (b) shows that our method can still effectively track the human body even if it is almost completely occluded by another player.

Next, we test the adaptation of our method to appearance changes. In Fig. 9, a particularly challenging (with high clutter) video sequence is used. The head of a player is tracked even though it moves fast and the appearance of the head changes frequently (including occlusion, blurring, and changes in the spatial and color distributions of the appearance). Fig. 9 shows that our method has successfully tracked the target and adapted to the changes of the target appearance.

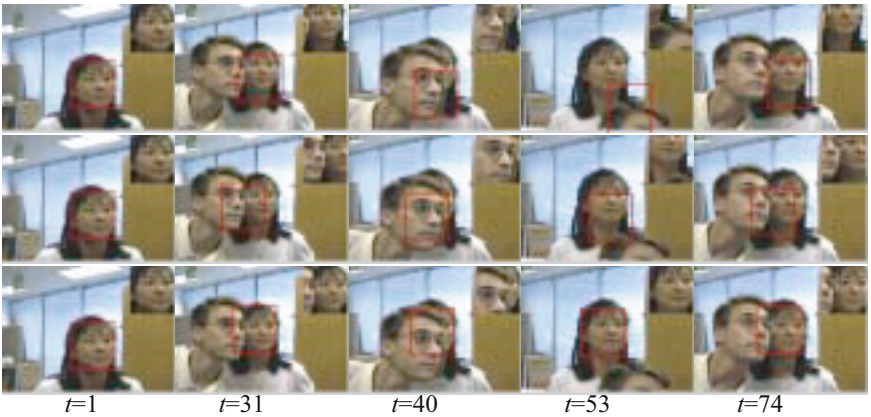


Fig. 7. Tracking results of the *girl* sequence with the MS tracker (first row), Condensation (second row) and our method (third row). The tracked face is shown in the upper-right window.

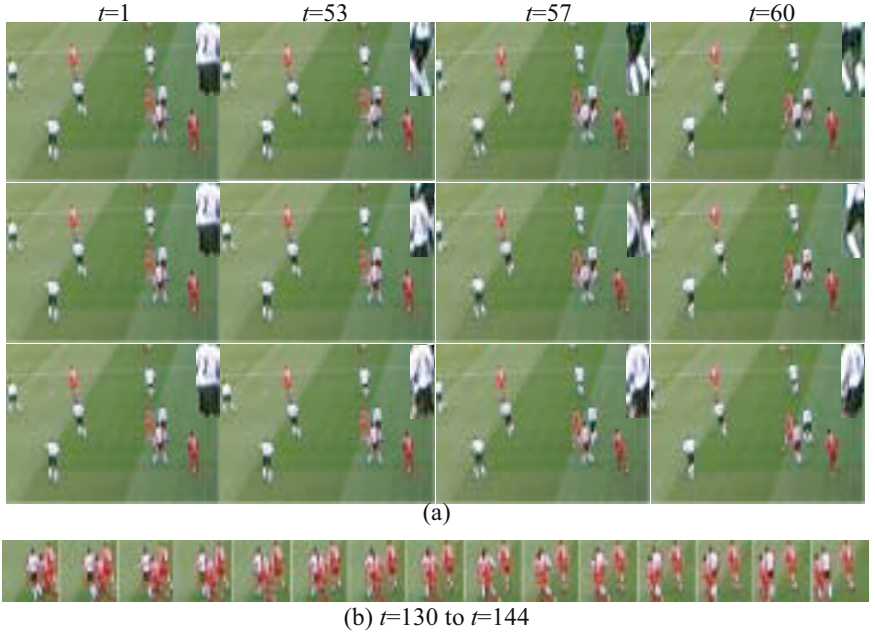


Fig. 8. Tracking results of the *soccer* sequence with three methods (a): the MS tracker (first row), Condensation (second row) and our method (third row). The tracked body is also shown in the upper-right window; (b) tracking results with occlusions by our method.

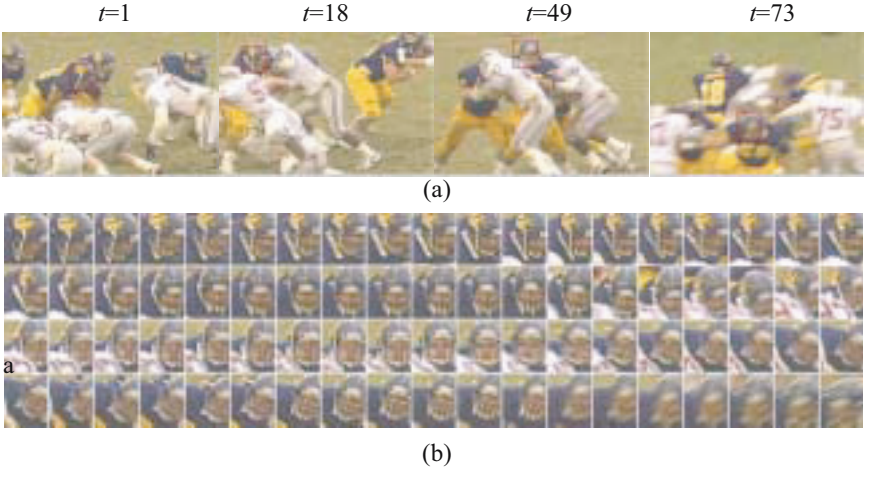


Fig. 9. (a) Tracking results of the *football* sequence with the MS tracker (first row), Condensation (second row) and our method (third row); (b) the target appearance changes (frames from 2-77)

5 Conclusion

We have described an effective appearance model (SMOG) in a joint spatial-color space and a new similarity measure based on SMOG. The SMOG appearance model and the SMOG-based similarity measure consider both the spatial distribution and the color distribution of objects: they utilize richer information than the general color histogram based appearance model and similarity measure.

We also propose an Integral Gaussian Mixture (IGM) technique, which greatly improves the computational efficiency of our method. Thus the number of particles and the size of target candidate region can be greatly increased, without significant change in the processing time of the proposed method.

We have successfully applied the SMOG appearance model and the SMOG-based similarity measure to the task of visual tracking in the framework of particle filters. Our tracking method can effectively handle clutter, illumination changes, appearance changes, occlusions, etc. Comparisons show that our method outperforms popular methods such as the general color histogram based MS tracker and Condensation.

Acknowledgements

We thank Dr. Chunhua Shen for his valuable comments, and the ARC for support (grant DP0452416).

References

1. Comaniciu, D., V. Ramesh, and P. Meer, *Kernel-based Object Tracking.* IEEE Trans. Pattern Analysis and Machine Intelligence, 2003. **25**(5): p. 564 - 577.
2. Nummiaroa, K., E. Koller-Meierb, and L.V. Gool, *An Adaptive Color-Based Particle Filter.* Image and Vision Computing, 2003. **21**: p. 99-110.
3. Isard, M. and A. Blake, *Condensation-Conditional Density Propagation for Visual Tracking.* International Journal of Computer Vision, 1998. **29**(1): p. 5-28.
4. Perez, P., et al. *Color-Based Probabilistic Tracking*. European Conference on Computer Vision. 2002. p. 661-675.
5. Shen, C., A.v.d. Hengel, and A. Dick. *Probabilistic Multiple Cue Integration for Particle Filter Based Tracking*. International Conference on Digital Image Computing - Techniques and Applications. 2003. p. 309-408.
6. McKenna, S.J., et al., *Tracking Groups of People.* Computer Vision and Image Understanding, 2000. **80**: p. 42-56.
7. Pérez, P., J. Vermaak, and A. Blake, *Data Fusion for Visual Tracking with Particles.* Proceedings of the IEEE, 2004. **92**(3): p. 495-513.
8. Yang, C., R. Duraiswami, and L. Davis. *Fast Multiple Object Tracking via a Hierarchical Particle Filter*. International Conference on Computer Vision. 2005. p. 212-219.
9. Porikli, F. *Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces*. Computer Vision and Pattern Recognition. 2005. p. 829-836.
10. Zhou, S., R. Chellappa, and B. Moghaddam, *Visual Tracking and Recognition Using Appearance-Adaptive Models in Particle Filters.* IEEE Transactions on Image Processing, 2004. **11**: p. 1434-1456.

11. Lichtenauer, J., M. Reinders, and E. Hendriks. *Influence of the Observation Likelihood Function on Particle Filtering Performance in Tracking Applications*. IEEE International Conference on Automatic Face and Gesture Recognition. 2004. p. 767-772.
12. Isard, M. and A. Blake. *ICONDENSATION: Unifying Low-level and High-level Tracking in a Stochastic Framework*. European Conference on Computer Vision. 1998. p. 893-908.
13. Wren, C.R., et al., *Pfinder: real-time tracking of the human body.* IEEE Trans. Pattern Analysis and Machine Intelligence, 1997. **19**(7): p. 780-785.
14. Elgammal, A., et al., *Background and Foreground Modeling using Non-parametric Kernel Density Estimation for Visual Surveillance.* Proceedings of the IEEE, 2002. **90**(7): p. 1151-1163.
15. Yang, C., R. Duraiswami, and L.S. Davis. *Efficient Mean-Shift Tracking via a New Similarity Measure*. Computer Vision and Pattern Recognition. 2005. p. 176-183.
16. McKenna, S.J., Y. Raja, and S. Gong, *Tracking Colour Objects Using Adaptive Mixture Models.* Image and Vision Computing, 1999. **17**: p. 225-231.
17. Stauffer, C. and W.E.L. Grimson. *Adaptive Background Mixture Models for Real-time Tracking*. Computer Vision and Pattern Recognition. 1999. p. 246-252.
18. Han, B. and L. Davis. *On-Line Density-Based Appearance Modeling for Object Tracking*. International Conference on Computer Vision. 2005. p. 1492-1499.
19. Wu, Y. and T.S. Huang, *Robust Visual Tracking by Integrating Multiple Cues Based on Co-Inference Learning.* International Journal of Computer Vision, 2004. **58**(1): p. 55-71.
20. Khan, S. and M. Shah. *Tracking People in Presence of Occlusion*. Asian Conference on Computer Vision. 2000. p. 263-266.
21. Birchfield, S. and S. Rangarajan. *Spatiograms versus Histograms for Region-Based Tracking*. Computer Vision and Pattern Recognition. 2005. p. 1152-1157.
22. Viola, P. and M. Jones, *Robust Real-Time Face Detection.* International Journal of Computer Vision, 2004. **52**(2): p. 137-154.