Graphical Model based Cue Integration Strategy for Head Tracking

Xiaopin Zhong, Jianru Xue and Nanning Zheng
Institute of Artificial Intelligence and Robotics
Xi'an Jiaotong University, Xi'an, Shaanxi, China 710049
{xpzhong,jrxue,nnzheng}aiar.xjtu.edu.cn

Abstract

To achieve robust system, more and more vision researchers take into account fusing multiple visual cues. In this paper, we propose a novel strategy to integrate multiple naive cues for head tracking. Firstly, a cue dependency model is constructed via graphical model. Secondly, a new inference procedure based on non-parametric belief propagation is built for cue integration. The work presented is thus a general framework easy to extend for other computer vision research problems. Experimental results imply that the strategy we propose is effective, and it is robust without estimation of cue reliability.

1 Introduction

Computer vision systems, such as video surveillance, robot navigation, and intelligent vehicle, aim to understand the real world which is complex, changeful and unpredictable. Object tracking, whose purpose is to infer states of targets from observation in image sequences, is one of crucial communities in computer vision. A challenge in developing object tracking systems is the difficulty of tracking robustly and flexibly in real environment, where there are a lot of clutter, occlusion and multiple similar targets, furthermore the object itself is non-rigid or color-changing. Therefore, target representation is the key aspect impacting the performance of tracking systems.

In order to achieve robust perception, traditional researchers developed many ad-hoc target representations to enhance the performance of systems. In fact, however, most of them focus on single cue(or modality) which is based on specific assumptions. They are not able to make use of the visual information which is richly provided by natural scenes. Hence techniques of integrating multiple different cues to represent target are getting more attention from recent researchers to overcome the weaknesses and limitations of single cue. Cue integration makes use of complementary and redundant visual information, performing more reliably though the computation has increased.

Nevertheless, it is very difficult to fuse different cues to work because how to estimate the reliability of individual cue is another open topic [3]. In this paper, we present a novel graphical model based framework for visual cue integration in object tracking, omitting the estimation of cue reliability. There are three attributes in this framework: firstly, the integration procedure is really autonomous, and doesn't need any intervening; secondly, no estimation of cue reliability is needed any more; thirdly, it's indeed a framework easy to extend. In the following sections, the three attributes are discussed in detail.

2 Related Work

As stated above, more and more researchers are taking account of fusing multiple cues for robust machine vision. We now give a little overview on cue integration strategies. There are mainly three kinds of approaches: rule based methods [16, 13, 10], information theory based methods [5] and probabilistic approaches [8, 18, 1, 6].

Rule based methods use fixed rules to combine information, and most of them take the assumption of cue independence. It is supposed by Jacobs [3] that humans perceive scene through a weighted-sum rule mechanism and the results are controled by reliability of individual cue. Although an adaptive democratic integration for tracking via differential equation is introduced by Triesch and Malsburg [16], its result is still a weighted sum. A more general version of democratic integration is extended by Spengler and Schiele [13]. Moreover, *opinion pool* is used to build likelihood function of target by Shen [11]. In classification perspective, the relationship among most of rule-based combination strategies, such as product rule, sum rule and voting, is analyzed by Kittler [4]. Obviously, fixed rule in advance is the major drawback of rule based methods.

Information theory based cue fusion introduced by Kruppa [5] uses *mutual information* to combine object modalities hierarchically. Mutual information is also employed to represent the content of individual module in [9]. Mutual information of cues denotes a dependency measure among cues very well, however, the estimation of mutual information itself relies on joint distribution of all modules.

Probabilistic approaches describe individual cue by probability density function, with variance(or covariance) of density naturally implying uncertainty of modality. Hence they are most widely used in tracking literature. A general framework is proposed by Leichter [6] to combine different tracking algorithms, however, the modalities are assumed to be independent in advance. In [8] a product formulation of posterior densities is deduced via assuming sequentially conditionally dependent by Moreno-Noguer. Obviously, it's not a real dependent case either. *Jump Markov Chain* is employed by Wu [17] to switch several observation models for target representation and another multiple modalities fusion base on modality report and modality reliability is introduced by Toyama [15]. Furthermore, a co-inference method [18] by building a factorized graphical model is presented by Wu and Huang. When the number of modalities increases, however, it is not easy to set up algorithm for cue integration by [8] or [18] any more.

Graphical model is a marriage between probability theory and graph theory, playing an increasingly important role in computer vision. Although there are many techniques developed for making inference in graphical model, problems with high dimension in real world are still hard to deal with. Hua [1] successfully build a graphical model for multi-scale object tracking following the idea of non-parametric belief propagation [2, 14].

In the following sections, we propose a novel cue integration strategies via graphical model for object tracking. The rest of this paper is organized as follows: Section 3 demonstrates our model for cue integration. Thereinto, cue dependency model is introduced in subsection 3.1 and model inference method will be discussed in subsection 3.2. In section 4, cues for object tracking will be designed, and experimental results are given out and discussed. Finally, Section 5 concludes this paper.

3   Cue Dependency Model

As introduced in previous section, rule based cue integration method uses fixed rule, which is not a real robust technique. In addition, estimation of cue reliability is difficult. While graphical model is a power tool to deal with uncertainty and decouple dependency of patterns. In this section, we facilitate cue integration problem in a graphical model framework.

3.1   Basic Model

Suppose the object is represented by M modalities, each corresponding to a state space $\mathbf{S}_i$ and forming a joint space, i.e., Cartesian product space $\mathbf{S} = \mathbf{S}_1 \times \mathbf{S}_2 \times ... \times \mathbf{S}_M$. Let $\mathbf{x}_i \in \mathbf{S}_i$ be state of ith modality, with observation $\mathbf{z}_i$ respectively. At each time step, the state nodes of all cues are dependent on each other. Hence, they are fully connected to be a Gibbs random field(see Figure.1(a) for the generic graph).
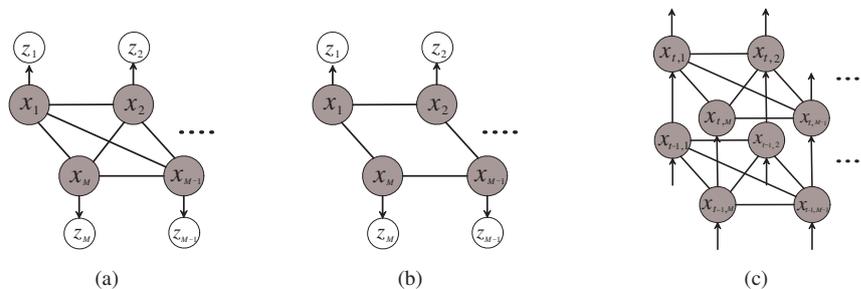

Figure 1: (a) Generic graph model for dependent cues, whose states form a Gibbs random field. The gray circles are hidden state nodes, and the white are observation nodes. (b) An example pruning some dependent paths to simplify the dependency model. (c) Model with dynamics, omitting the observation nodes.

In tracking literature, our goal is to make inference for joint density of all nodes in this graph. We denote joint state of target by $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M\} \in \mathbf{S}$ and joint observation by $\mathbf{z} = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_M\} \in \mathbf{O}$, where $\mathbf{x}_i$ is a variable in $\mathbf{S}_i$ and $\mathbf{O}$ is observation space. For simplification, the model is restricted to pairwise Markov random network. Then the joint density of state nodes and observation nodes can be formulated as follows,

$$P(\mathbf{x}, \mathbf{z}) = \kappa \prod_{i=1}^{M} \phi_i(\mathbf{x}_i, \mathbf{z}_i) \prod_{j \in \Gamma(i)} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j), \tag{1}$$

where ϕ_i and ψ_{ij} are the potential functions of single node clique and pair node clique respectively, with κ being a normalizing constant and $\Gamma(i)$ neighborhood set of node i.

Generally, ϕ_i is taken as a measurement likelihood function which is differently produced by each cue in our case. On the other hand, the interaction between two neighbor nodes of different cues is too complicated to formulate. Therefore, we make a mixture of Gaussians assumption for potential of two neighbor nodes, i.e. following the idea in [12]

$$\psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \pi_0 \mathcal{N}(\mathbf{x}_j; \mu_{ij}, \Lambda_{ij}) + (1 - \pi_0) \sum_{m=1}^{M_{ij}} \pi_m \mathcal{N}(\mathbf{x}_j; \mu_{i,m}, \Sigma_{i,m}), \tag{2}$$

where $\sum_{m=1}^{M_{ij}} \pi_m = 1$. $\mathscr{N}(\mathbf{x}_j; \mu_{ij}, \Lambda_{ij})$ is a outlier process and $\mathscr{N}(\mathbf{x}_j; \mu_{i,m}, \Sigma_{i,m})$ is a normal distribution with mean $\mu_{i,m}$ and covariance $\Sigma_{i,m}$. Obviously, Σ_i is covariance matrix of ith cue to represent its uncertainty. This approximation is indeed reasonable since it makes the cues to be evidence for each other. For simplicity, we use $M_{ij} = 1$ in our experiments. Particle framework is adopted in our inference strategy, therefore the parameters can be estimated by EM algorithm on weighted sample set.

In fact, if the number of modalities is large, pruning should be taken into account to avoid getting into problem intractable (see Fig.1(b)). Base on this model, our cue integration procedure seems like shrinking process of an elastic curve. During model inference, the cue estimations converge because they will gather evidence from each other via passing message in a cyclic graph. The evidence power is internal force, whilst observation is external force. Therefore, the elastic curve will stop shrinking when the two forces go to a fine balance, and that is the exact time of accomplishment of model inference.

In view of tracking task, an overall target estimation is needed. To make it clearer, we assume that overall state space and the state space of individual modality are identical, say $\mathbf{S}_o = \mathbf{S}_1 = \mathbf{S}_2 = ... = \mathbf{S}_M$. Even if this does not hold, certain task dependent space mapping function could be defined to help. After model inference process, obviously, all the weak cues are enhanced mutually to be independent on each other. Therefore, all traditional fusion methods based on independence assumption are reasonable choices for overall estimation. Product rule is preferred in our experiments for simplicity, i.e. $P(\mathbf{x}_o) \propto \prod_i P(\mathbf{x}_i)$, where $\mathbf{x}_o \in \mathbf{S}_o$.

3.2 Model Inference

After constructing a graphical model for cue integration, it remains to make inference for marginal probability distribution of each modality. If variables in the model are discrete, all potentials are easily achieved. However, High dimension problem in real situation challenges exhaustive discretization of state space a lot. Fortunately, Monte Carlo technique makes it possible to make approximate inference by performing numeric computation. Many of such algorithms [14, 2, 1] have been developed. And all of them have argued that non-parametric belief propagation could make inference even in graph with circles. We adopt the algorithm introduced in [19] for the reason of sequential demand in tracking.

In Bayesian tracking perspective, posterior of individual cue is the key:

$$p(\mathbf{x}_i|\mathbf{z}) \propto \int \cdots \int p(\mathbf{x},\mathbf{z})d\mathbf{x}_{-i} \propto \int \cdots \int \prod_{i=1}^{M} \phi_i(\mathbf{x}_i,\mathbf{z}_i) \prod_{j \in \Gamma(i)} \psi_{ij}(\mathbf{x}_i,\mathbf{x}_j)d\mathbf{x}_{-i}, \qquad (3)$$

where $\mathbf{x}_{-i} \triangleq \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, ..., \mathbf{x}_M\}$. To treat this N-P hard problem, BP(Belief Propagation) algorithm employs an evidence message from node i to node j as follows,

$$m_{ij}^k(\mathbf{x}_j) = \int \phi_i(\mathbf{x}_i,\mathbf{z}_i)\psi_{ij}(\mathbf{x}_i,\mathbf{x}_j) \prod_{u \in \Gamma(i)\backslash j} m_{ui}^{k-1}(\mathbf{x}_i)d\mathbf{x}_i, \qquad (4)$$

where k is used to indicate message updating recursion. Finally, BP approximates the belief by

$$p(\mathbf{x}_i|\mathbf{z}) \propto \phi_i(\mathbf{x}_i,\mathbf{z}_i) \prod_{j \in \Gamma(i)} m_{ji}^k(\mathbf{x}_i), \qquad (5)$$

after k iterations on message updating.

In view of non-parametric belief propagation, all messages and beliefs are represented by weighted sample set. With the help of particle filtering technique who uses importance sampling to propagate belief, we could easily initialize message in (4). However, it's difficult to find a suitable importance function because message itself is hard to describe even using mixture of Gaussians. Therefore, sampling procedure is stratified by a parameter v, $0 \leq v \leq 1$, to enhance the strength of sample set. Finally, a modified non-parametric belief propagation is introduced(see Figure.2 for one iteration).

Considering object dynamics(see Figure.3(c) for dynamic Markov network), the message and belief are rewritten by adding temporal messages, say

$$
m_{ij}^k(\mathbf{x}_{t,j}) = \int \phi_i(\mathbf{x}_{t,i}, \mathbf{z}_{t,i}) \psi_{ij}(\mathbf{x}_{t,i}, \mathbf{x}_{t,j}) \times
$$

$$
\int p(\mathbf{x}_{t,i}|\mathbf{x}_{t-1,i}) P(\mathbf{x}_{t-1,i}|\mathbf{z}_{t-1}) d\mathbf{x}_{t-1,i} \prod_{u \in \Gamma(i) \setminus j} m_{ui}^{k-1}(\mathbf{x}_{t,i}) d\mathbf{x}_{t,i}, \tag{6}
$$

$$
p(\mathbf{x}_{t,i}|\mathbf{z}_t) \propto \phi_i(\mathbf{x}_{t,i}, \mathbf{z}_{t,i}) \prod_{j \in \Gamma(i)} m_{ji}^k(\mathbf{x}_{t,i}) \cdot \int p(\mathbf{x}_{t,i}|\mathbf{x}_{t-1,i}) p(\mathbf{x}_{t-1,i}|\mathbf{z}_{t-1}) d\mathbf{x}_{t-1,i}. \tag{7}
$$

According to algorithm in Figure.2, we could easily accomplish the entire sequential version of inference process by substituting (6),(7) for message and belief.
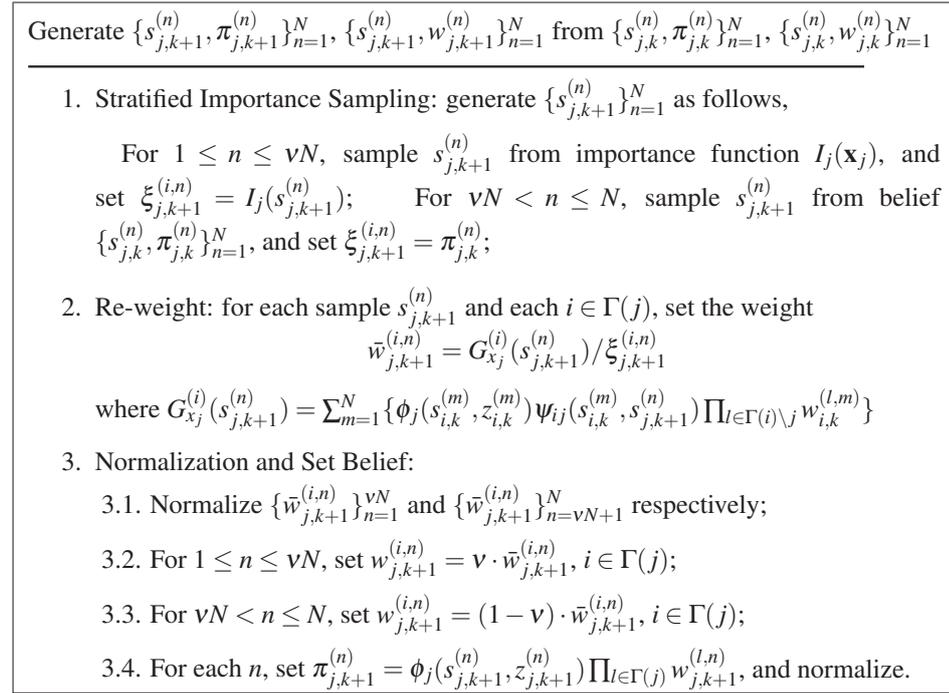

Generate $\{s_{j,k+1}^{(n)}, \pi_{j,k+1}^{(n)}\}_{n=1}^N$, $\{s_{j,k+1}^{(n)}, w_{j,k+1}^{(n)}\}_{n=1}^N$ from $\{s_{j,k}^{(n)}, \pi_{j,k}^{(n)}\}_{n=1}^N$, $\{s_{j,k}^{(n)}, w_{j,k}^{(n)}\}_{n=1}^N$

1. Stratified Importance Sampling: generate $\{s_{j,k+1}^{(n)}\}_{n=1}^N$ as follows,

   For $1 \leq n \leq vN$, sample $s_{j,k+1}^{(n)}$ from importance function $I_j(\mathbf{x}_j)$, and set $\xi_{j,k+1}^{(i,n)} = I_j(s_{j,k+1}^{(n)})$; For $vN < n \leq N$, sample $s_{j,k+1}^{(n)}$ from belief $\{s_{j,k}^{(n)}, \pi_{j,k}^{(n)}\}_{n=1}^N$, and set $\xi_{j,k+1}^{(i,n)} = \pi_{j,k}^{(n)}$;

2. Re-weight: for each sample $s_{j,k+1}^{(n)}$ and each $i \in \Gamma(j)$, set the weight
   $$
   \bar{w}_{j,k+1}^{(i,n)} = G_{x_j}^{(i)}(s_{j,k+1}^{(n)})/\xi_{j,k+1}^{(i,n)}
   $$
   where $G_{x_j}^{(i)}(s_{j,k+1}^{(n)}) = \sum_{m=1}^N \{\phi_j(s_{i,k}^{(m)}, z_{i,k}^{(m)}) \psi_{ij}(s_{i,k}^{(m)}, s_{j,k+1}^{(n)}) \prod_{l \in \Gamma(i) \setminus j} w_{i,k}^{(l,m)}\}$

3. Normalization and Set Belief:

   3.1. Normalize $\{\bar{w}_{j,k+1}^{(i,n)}\}_{n=1}^{vN}$ and $\{\bar{w}_{j,k+1}^{(i,n)}\}_{n=vN+1}^N$ respectively;

   3.2. For $1 \leq n \leq vN$, set $w_{j,k+1}^{(i,n)} = v \cdot \bar{w}_{j,k+1}^{(i,n)}$, $i \in \Gamma(j)$;

   3.3. For $vN < n \leq N$, set $w_{j,k+1}^{(i,n)} = (1-v) \cdot \bar{w}_{j,k+1}^{(i,n)}$, $i \in \Gamma(j)$;

   3.4. For each n, set $\pi_{j,k+1}^{(n)} = \phi_j(s_{j,k+1}^{(n)}, z_{j,k+1}^{(n)}) \prod_{l \in \Gamma(j)} w_{j,k+1}^{(l,n)}$, and normalize.

Figure 2: One iteration of modified Non-parametric Belief Propagation

4 Experiments

In this section, a head tracking problem is tackled in our cue integration framework. We define three weak cues and their observation functions in advance, and then discuss the

experiment results.

4.1 Cues Selected for Robust Head Tracking

In this subsection, we select three naive cues and define their likelihood functions for tracking heads in real environments. They are color, ellipse shape and intensity change. For simplicity, the head's projection on image is approximated by an ellipse. Furthermore, in order to make our cue integration approach clear, we set all cues in the same state spaces which is a four-D upright ellipse space. Hence, we model the head configuration as an ellipse parameterized by $\mathbf{x}_i = (x_0, y_0, a, b)$, $i = 1, ..., M$, where coordinate (x_0, y_0) is the center of ellipse with a and b being the lengths of two axes respectively. It is noted that all the states have dropped the time slice subscript in this subsection.

4.1.1 Color Cue

Color has been proved to be a simple but effective feature in computer vision. Due to statistical characteristic, MoG(Mixture of Gaussians) is often used to model color distribution and easy to trained via EM(Expectation-Maximization) algorithm. Therefore, MoG is employed to model head color which mainly consists of face(skin) color and hair color. Then given parameter Θ, the probability of the color $\mathbf{c}$ belonging to head is formulated as $p_c(\mathbf{c}|head, \Theta) = \sum_{i=1}^{M_c} \alpha_{c,i} \mathcal{N}(\mathbf{c}; \mu_{c,i}, \Sigma_{c,i})$, where $\Theta = \{\alpha_{c,i}, \mu_{c,i}, \Sigma_{c,i}\}_{i=1}^{M_c}$ is EM-trained parameter set with the number of Gaussians M_c.
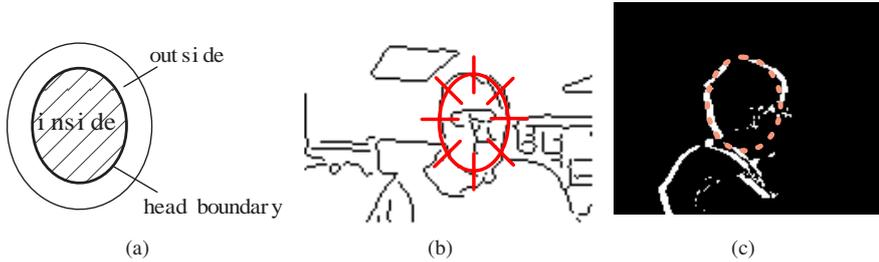

(a)   (b)   (c)

Figure 3: (a)Two concentric elliptical areas to locate the head. Elliptical area *inside* indicates the head area, and Elliptical area *outside* is the transition area from head to background. (b)Measurement line of ellipse shape cue. (c)Measurement point of intensity-change cue.

Further we model the likelihood of observing an image given a head configuration $\mathbf{x}_i$ by the following way. Firstly, the area inside head and the area outside head are defined along the boundary of head as Fig.3(a). Then we calculate the average foreground probability of inside area as,

$$f_{in}(\mathbf{x}_i) = \frac{1}{N_{in}} \sum_{k=1}^{N_{in}} p_{c_k}(\mathbf{c}_k|head, \Theta) \tag{8}$$

where N_{in} is the number of pixels in inside area and $\mathbf{c}_k$ is the color value of *kth* pixel. The average foreground probability of outside area is similarly defined as $f_{out}(\mathbf{x}_i) = \frac{1}{N_{out}} \sum_{k=1}^{N_{out}} p_{c_k}(\mathbf{c}_k|head, \Theta)$.

Finally, a likelihood function for head configuration is yielded,

$$p_{cl}(\mathbf{z}_i|\mathbf{x}_i) \propto \exp\{-\lambda_{c,1} \cdot \frac{(f_{in})^{\lambda_{c,2}}}{f_{out}}\}, \tag{9}$$

where $\lambda_{c,1}$ is used for adjusting the sensitivity to head configuration and $\lambda_{c,2}$ for tuning the impact of outside area.

4.1.2 Ellipse Shape Cue

The second simple cue we choose is ellipse shape cue, following the model introduced by MacCormick [7]. Unlike their complicated B-Spline contour model, we draw N_{nl} measurement lines normal to elliptical head contour directly (see Figure.3(b)).

An edge detector is applied on each measurement line, then we have $n_l(l = 1,...,N_{nl})$ edge points used for features on the *lth* measurement line. The distance between the *jth* feature and ellipse contour is denoted by z_j^l, $j = 1,...,n_l$. In addition, this distance is assumed to be Gaussian, with zero mean and variance σ_s^2. Let the probability of no feature detected on measurement line be p_0 and the number of feature generated by clutter obey Poisson process with density λ_s. Then, the likelihood of observation on *lth* measurement line is defined as

$$\mathcal{L}_s(z^l|\mathbf{x}_i) \propto p_0 + \frac{1-p_0}{\sqrt{2\pi}\sigma_s\lambda_s} \sum_{j=1}^{n_l} exp\{-\frac{(z_j^l)^2}{2\sigma_s^2}\}. \tag{10}$$

Consequently, we take the independent assumption of likelihoods on each measurement line. Thus the overall ellipse shape likelihood is

$$p_{sh}(\mathbf{z}_i|\mathbf{x}_i) = \prod_{l=1}^{N_{nl}} \mathcal{L}_s(z^l|\mathbf{x}_i). \tag{11}$$

4.1.3 Intensity-Change

The last weak cue we use in head tracking is intensity-change, which is based on difference of neighbor frames. With no doubt, intensity-change provide us target's motion cue because we always assume the moving of camera is trivial.

A new likelihood function is introduced for ellipse state as follows. Firstly, the motion area(white area in Fig.3(c)) is obtained by thresholding the difference of neighbor frames. Then we put a lot of points on ellipse boundary with equidistance, named measure points, and count the number of measure points that have been covered by motion area $N_{e,co}$, and count the number of pixels covered by motion area inside the ellipse $N_{p,co}$. Further, we denote the total number of measure points by $N_{e,to}$ and the total number of pixels inside ellipse by $N_{p,to}$. We could define likelihood as $exp\{-\lambda_{in,1}\frac{N_{e,co}}{N_{e,to}}\}$ in intuition, however pixels inside object area should not be motion area owing to slight motion. Thus, a penalty is needed, then the likelihood function is achieved as

$$p_{in}(\mathbf{z}_i|\mathbf{x}_i) \propto \exp\{-\lambda_{in,1}(\frac{N_{e,co}}{N_{e,to}} - \lambda_{in,2}\frac{N_{p,co}}{N_{p,to}})\}, \tag{12}$$

where parameter $\lambda_{in,1}$ adjusts the sensitivity to intensity-change cue and $\lambda_{in,2}$ adjusts the penalty of slight motion.

As emphasized before, we preliminarily focus on how to integrate several cues to get robust result, but not how to develop effective cues. Therefore the three weak cues stated above for rough object representation are selected.

4.2 Experimental Results

In this subsection, a group of experimental results[1] are shown. In all of the figures below, we use red ellipse to indicate MMSE estimation of color cue, and green for shape cue, yellow for intensity-change cue.

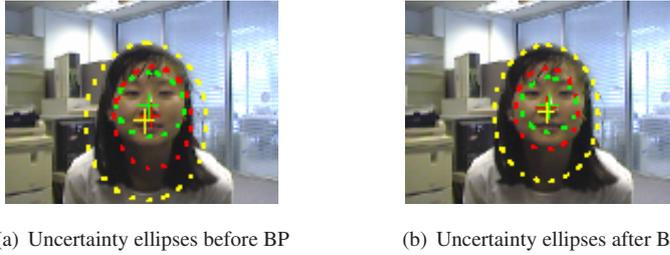(a) Uncertainty ellipses before BP   (b) Uncertainty ellipses after BP

Figure 4: Position uncertainty ellipses before and after Belief Propagation.

Figure.4 illustrates median results before and after belief propagation. The sign of + denotes the position estimation of individual cue, whilst the corresponding ellipses show their uncertainty only on position. It is easy to found that after belief propagation, all of the cues become more reliable[2] and the estimations on the three cues become close but still distinct from each other. It is exactly proved that our integration procedure seems like shrinking of an elastic curve[3].
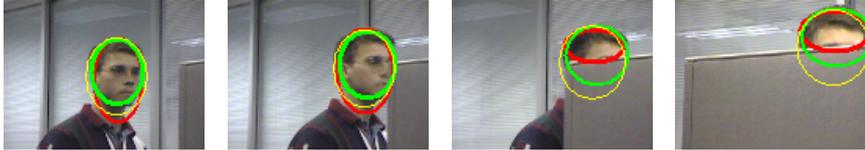

Figure 5: Head tracking results in sequence 'seq_cubicle.avi'. From left to right, the corresponding frame number is 52, 56, 60 and 68. The thickness of ellipse represents the reliability of each cue.

Figure.5 shows the results of head tracking with occlusion in sequence 'seq_cubicle.avi'. And we use the thickness of ellipse to represent the reliability[4], determined by uncertainty of each cue. Obviously, the reliability is outgrowth of our tracking procedure. Thus, different from other strategies, graphical model based cue integration doesn't rely on any cue reliability in advance. Moreover, uncertainties of all cues decrease, implying that our integration strategy really makes tracking system more reliable.

Some results of head tracking in sequence 'seq_sb.avi' are available in Figure.6 for performance comparison. When tracking without(top row) integration, color, shape and intensity-change fail in frame 123, 164 and 133 respectively. It's easily concluded that

---

[1] The test sequences are available at http://robotics.stanford.edu/ birch/headtracker/seq/.

[2] Covariance or uncertainty has decreased.

[3] It has been discussed in subsection 3.1.

[4] An alternative of reliability here is reciprocal of L_2 norm of particles covariance.
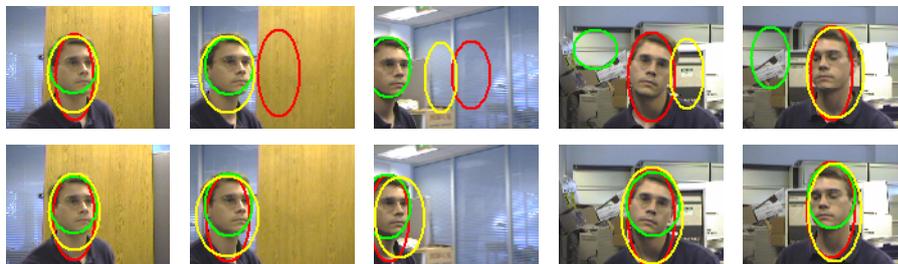
Figure 6: Head tracking results in sequence 'seq_sb.avi'. The top row results are produced by three independent cues without integration. In contrast, the bottom row are produced by our cue integration method. From left to right, the corresponding frame number is 116, 123, 133, 164 and 183.

other integration strategies base on independence assumption will fail similarly, even reliability can help to estimate the overall result. In contrast, our integration strategy is able to track the head all the time(bottom row) robustly and flexibly.

5 Conclusion

This paper has addressed the important problem of cue integration in object tracking. And a novel cue integration strategy is proposed base on graphical model theory. The properties of our strategy have been demonstrated as follows,

- a cue dependency model is built via graphical model;
- the integration process is autonomous, attributing to new inference method base on non-parametric belief propagation;
- it's indeed easy to extend, without any redundant estimation of cue reliability.

In future work, we intend to develop efficient computation techniques for our integration strategy, and apply it to more research problems in computer vision.

Acknowledgement

References

[1] Gang Hua and Ying Wu. Multi-scale visual tracking by sequential belief propagation. In *Proceedings of CVPR*, pages I:826–833, 2004.

[2] Michael Isard. Pampas: Real-valued graphical models for computer vision. In *Proceedings of CVPR*, pages I:613–620, 2003.

[3] Robert A. Jacobs. What determines visual cue reliability? *TRENDS in Cognitive Sciences*, 6(8):345–310, 2002.

[4] Josef Kittler, Mohamad Hatef, and Robert P.W. Duin. On combining classifiers. *IEEE Trans. on PAMI*, 20(3):226–239, 1998.

[5] Hannes Kruppa and Bernt Schiele. Hierarchical combination of object models using mutual information. In *Proceedings of 12th BMVC*, pages 103–112, 2001.

[6] Ido Leichter, Michael Lindenbaum, and Ehud Rivlin. A probabilistic framework for combining tracking algorithms. In *Proceedings of CVPR*, pages II:445–451, 2004.

[7] John MacCormick and Andrew Blake. Probabilistic exclusion and partitioned sampling for multiple object tracking. *IJCV*, 39(1):57–71, 2000.

[8] Francesc Moreno-Noguer, Alberto Sanfeliu, and Dimitris Samaras. Integration of conditionally dependent object features for robust figure/background segmentation. In *Proceedings of ICCV*, pages 1713–1720, 2005.

[9] Joseph L. Mundy and Chung-Fu Chang. Fusion of intensity, texture, and color in video tracking based on mutual information. In *Proceedings of the 33rd Applied Imagery Pattern Recognition Workshop (AIPR'04)*, pages 10–15, 2004.

[10] Ipek Oruç, Laurence T. Maloney, and Michael S. Landy. Weighted linear cue combination with possibly correlated error. *Vision Research*, 43:2451–2468, 2003.

[11] Chunhua Shen, Anton van den Hengel, and Anthony Dick. Probabilistic multiple cue integration for particle filter based tracking. In *Proceedings of 7th International Conference on Digital Image Computing: Techniques and Applications*, pages I:399–408, 2003.

[12] Leonid Sigal, Sidharth Bhatia, Michael J.Black, and Michael Isard. Tracking loose-limbed people. In *Proceedings of the CVPR*, pages I:421–428, 2004.

[13] Martin Spengler and Bernt Schiele. Towards robust multi-cue integration for visual tracking. *Machine Vision and Applications*, 14:50–58, 2003.

[14] Erik B. Sudderth, Alexander T. Ihler, William T. Freeman, and Alan S. Willsky. Nonparametric belief propagation. In *Proceedings of CVPR*, pages I:605–612, 2003.

[15] Kentaro Toyama and Eric Horvitz. Bayesian modality fusion: Probabilistic integration of multiple vision algorithms for head tracking. In *Proceedings of Asian Conference on Computer Vision*, 2000.

[16] Jochen Triesch and Christoph von der Malsburg. Democratic integration: Self-organized integration of adaptive cues. *Neural Computation*, 13:2049–2074, 2001.

[17] Ying Wu, Gang Hua, and Ting Yu. Switching observation models for contour tracking in clutter. In *Proceedings of CVPR*, pages I:295–302, 2003.

[18] Ying Wu and Thomas S. Huang. Robust visual tracking by integrating multiple cues based on co-inference learning. *IJCV*, 58(1):55–71, 2004.

[19] Jianru Xue, Nanning Zheng, and Xiaopin Zhong. Sequential stratified sampling belief propagation for multiple targets tracking. *Science in China: Series F Information Sciences*, 49(1):48–62, 2006.