Measurement integration under inconsistency for robust tracking

Gang Hua, Ying Wu
Department of Electrical Engineering and Computer Science
Northwestern University, 2145 Sheridan Road, Evanston, IL 60208
{ganghua, yingwu}@ece.northwestern.edu

Abstract

The solutions to many vision problems involve integrating measurements from multiple sources. Most existing methods rely on a hidden assumption, i.e., these measurements are consistent. In reality, unfortunately, this may not hold. The fact that naively fusing inconsistent measurements amounts to failing these methods indicates that this is not a trivial problem. This paper presents a novel approach to handling it. A new theorem is proven that gives two algebraic criteria to examine the consistency and inconsistency. In addition, a more general criterion is presented. Based on the theoretical analysis, a new information integration method is proposed and leads to encouraging results when applied to the task of visual tracking.

1 Introduction

In many vision problems, estimations are made based on integrating measurements from multiple sources to reduce the uncertainty. A measurement can generally be characterized as a mean vector and an uncertainty covariance (multi-modal measurement can be treated as multiple measurements). To list a few examples, the different sources can be different visual cues such as color and contour [14, 11], different components of one object [13, 10, 3, 4], neighborhood pixels in motion estimation [1], and dynamics and image observations in visual tracking [5].

Most existing integration methods assume the consistency among various sources [6, 7]. If the different sources are independent and consistent, the optimal integration can be obtained from the best linear unbiased estimator (BLUE) [7]. If they are correlated but consistent, the covariance intersection (CI) [6] obtains a consistent and conservative estimate. However, the consistency assumption may not hold in practice. In principle, if two measurements can be regarded as being generated from the same model (e.g., a Gaussian), then they are consistent. Otherwise they are inconsistent. The measurements from different sources can be very confident (i.e., small covariance) but are quite different. They do not agree with one another and it makes less sense to fuse them together forcefully. Measurement inconsistency fails both the BLUE and CI.

Indeed, this problem is not uncommon in computer vision applications. For example, a wrong dynamic prediction in Bayesian visual tracking is very likely to be inconsistent with the detected image observations. This is especially true when the target presents sudden dynamic changes. Such kind of inconsistency shall fail Kalman filtering that is based on BLUE. In part-based tracking, the measurements of different parts may be conflicting when some parts are distracted by camouflages. The aperture problem in motion estimation is another example [1].

Unfortunately, the handling of inconsistency is not well addressed in the literature. Therefore, it is desirable to carry out some basic study of inconsistency in order to identify the solution to robust measurement integration. We are particularly interested in answering two questions: (a) how can we detect inconsistency from the measurements? And (b) how can we handle it in integration? We need to develop principled criteria to characterize inconsistency and develop efficient method to detect and resolve it.

This paper describes a novel *distributed* integration approach based on the theory of Markov networks. Although Markov networks were widely applied to solve visual inference problems [2, 10, 13], the study of information fusion of the inference over Markov networks is largely remained unexplored. We proved a new theorem that provides two algebraic criteria to examine the *consistency* and *inconsistency* for pair-wise measurements. In addition a general criterion is proposed to detect inconsistency in a general setting.

Since the presence of inconsistency implies the presence of false or outlier measurements, our method can automatically identify the inconsistent measurements and eliminate the false ones for further integration. Based on the proposed integration approach, we have developed a robust part-based tracking algorithm in which measurements of various parts are robustly integrated for tracking, even when there exists some inconsistent ones.

There are some previous works that were aware of the *inconsistency* problem such as the covariance union (CU) [12] and the variable bandwidth density fusion (VBDF) [1]. They either increase the covariance of the integrated estimate to achieve covariance consistency with each of the integrated measurements [12], or seek for the most salient mode across all scales of the measurements kernel density [1]. None of them provides a principled criterion to evaluate measurement inconsistency, i.e., they are not able to determine when two measurements can be regarded as being obtained from one model.

2 Formulation of multi-source integration

Markov network provides a principled methodology for the *distributed* integration of multiple sources. The joint posterior defined on a Markov network is

$$p(\mathbf{X}|\mathbf{Z}) = \frac{1}{C}\prod_{\{i,j\}\in\mathcal{E}}\psi(\mathbf{x}_i,\mathbf{x}_j)\prod_{i\in\mathcal{V}}\phi(\mathbf{x}_i,\mathbf{z}_i), \quad (1)$$

where C is a normalization constant, $\mathbf{X} = \{\mathbf{x}_i : i = 1\dots N\}$, $\mathbf{Z} = \{\mathbf{z}_i : i = 1\dots N\}$ and N is the number of sources modeled in the Markov network.

Each $\mathbf{x}_i$ denotes the integrated estimate at node i, and $\mathbf{z}_i$ is the local measurement of source i. Set $\mathcal{V}$ indicates the set of $\{\mathbf{x}_i, \mathbf{z}_i\}$ pairs and each pair has a compatibility function $\phi(\mathbf{x}_i, \mathbf{z}_i)$. Let $\mathbf{x}_i$, $\mathbf{z}_i$ be in $\mathcal{R}^n$, since the measurement is a $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$ pair, $\phi(\mathbf{x}_i, \mathbf{z}_i)$ is in nature a Gaussian, i.e.,

$$\phi(\mathbf{x}_i,\mathbf{z}_i) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{\Sigma}_i|}}\mathbf{e}^{\left(-\frac{1}{2}(\mathbf{z}_i-\mathbf{x}_i)^T\mathbf{\Sigma}_i^{-1}(\mathbf{z}_i-\mathbf{x}_i)\right)}. \quad (2)$$

Set $\mathcal{E}$ defines the neighborhood relationships in the Markov network. If $\mathbf{x}_j$ is the neighbor of $\mathbf{x}_i$, then $\mathbf{x}_j$ can provide a predictive estimate $f_{ij}(\mathbf{x}_j)$ for $\mathbf{x}_i$. $\psi(\mathbf{x}_i, \mathbf{x}_j)$ is the compatibility function of the neighboring $\mathbf{x}_i$ and $\mathbf{x}_j$, i.e., a Gaussian

$$\psi(\mathbf{x}_i,\mathbf{x}_j) = \frac{\exp\left\{-\frac{(\mathbf{x}_i-f_{ij}(\mathbf{x}_j))^T(\mathbf{x}_i-f_{ij}(\mathbf{x}_j))}{2\sigma_{ij}^2}\right\}}{\sqrt{(2\pi)^n}\sigma_{ij}^n} \quad (3)$$

$$\doteq \frac{\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{A}_{ij}\mathbf{x}_j-\mu_{ij})^T(\mathbf{x}_i-\mathbf{A}_{ij}\mathbf{x}_j-\mu_{ij})}{2\sigma_{ij}^2}\right\}}{\sqrt{(2\pi)^n}\sigma_{ij}^n}, \quad (4)$$

which indicates if $\mathbf{x}_i$ and $f_{ij}(\mathbf{x}_j)$ can be regarded as being drawn from one common model and σ_{ij}^2 is the scalar variance. When f_{ij} is nonlinear, we linearize it by Taylor expansion, i.e., $\mu_{ij} = f_{ij}(\mathbf{0})$ and $\mathbf{A}_{ij} = \frac{\partial f_{ij}(\mathbf{x}_j)}{\partial \mathbf{x}_j}|_{\mathbf{x}_j=\mathbf{0}}$ is the $n \times n$ Jacobian. So we only consider the setting of Eq. 4. The σ_{ij}^2 indeed models the uncertainties between the local estimate $\mathbf{x}_i$ and the neighborhood estimate $\mathbf{A}_{ij}\mathbf{x}_j + \mu_{ij}$.

The integration of all the measurements is to perform the Bayesian inference on Eq. 1. Nevertheless, when some measurements are inconsistent with the others, it indicates there are false ones. Blindly integrating them will jeopardize the whole integration process. Let $\mathbf{O} = \{\mathbf{O}_i, i = 1\dots N\}$ be the binary set to indicate if $\mathbf{z}_i$ is false, i.e., $\mathbf{O}_i = 1$ means it is and vice versa. $\mathbf{O}$ divides $\mathbf{Z}$ into two sets, i.e., the false set $\mathbf{Z}_{\mathcal{O}}$ and the normal set $\mathbf{Z}_{\bar{\mathcal{O}}} = \mathbf{Z} \setminus \mathbf{Z}_{\mathcal{O}}$. Reliable integration requires eliminating the false ones, i.e., we should perform the Bayesian inference on

$$p(\mathbf{X}|\mathbf{Z}_{\bar{\mathcal{O}}}) = \frac{1}{C'}\prod_{\{i,j\}\in\mathcal{E}}\psi(\mathbf{x}_i,\mathbf{x}_j)\prod_{\mathbf{z}_i\in\mathbf{Z}_{\bar{\mathcal{O}}}}\phi(\mathbf{x}_i,\mathbf{z}_i), \quad (5)$$

where C' is again for normalization. Before we can achieve that, we need a rigorously criteria to judge *inconsistency*. For integration, this concept is always qualitative [12], we proceed to provide principled quantitative criteria.

3 Measurements inconsistency

Intuitively, assume $\mathbf{A}_{ij}$ and μ_{ij} be known, given all the $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$, the estimate of σ_{ij}^2 is a natural indicator of whether $\mathbf{x}_i$ and $\mathbf{A}_{ij}\mathbf{x}_j + \mu_{ij}$ is consensus, i.e., if σ_{ij}^2 is very small, then they are consensus since $\psi(\mathbf{x}_i, \mathbf{x}_j)$ is approaching to a delta function, and vice versa. Denote $\mathbf{\Theta} = \{\sigma_{ij}^2 : \{i,j\} \in \mathcal{E}\}$, Eq. 1 is indeed $p(\mathbf{X}|\mathbf{\Theta}, \mathbf{Z})$. The MAP estimate of $\mathbf{x}_i$ and the ML estimate of $\mathbf{\Theta}$ can be obtained by the following Bayesian EM algorithm [8], i.e.,

$$\begin{aligned}\mathbf{x}_i &= (\mathbf{\Sigma}_i^{-1} + \sum_{j\in\mathcal{N}(i)}\frac{1}{\sigma_{ij}^2}\mathbf{I})^{-1} \\ &\times (\mathbf{\Sigma}_i^{-1}\mathbf{z}_i + \sum_{j\in\mathcal{N}(i)}\frac{1}{\sigma_{ij}^2}(\mathbf{A}_{ij}\mathbf{x}_j + \mu_{ij})) \quad (6)\end{aligned}$$

$$\sigma_{ij}^2 = \frac{1}{n}(\mathbf{x}_i - \mathbf{A}_{ij}\mathbf{x}_j - \mu_{ij})^T(\mathbf{x}_i - \mathbf{A}_{ij}\mathbf{x}_j - \mu_{ij}) \quad (7)$$

Fixing $\mathbf{\Theta}$, the E-Step in Eq. 6 obtains the MAP estimate of $\mathbf{x}_i$ by fixed-point iteration. It is actually performing the BLUE [7] fusion of the local estimate and neighborhood estimate. Fixing $\mathbf{X}$, the M-Step in Eq. 7 maximizes $p(\mathbf{X}|\mathbf{\Theta}, \mathbf{Z})$ w.r.t. $\mathbf{\Theta}$. Combining the two steps together also constitutes a fixed-point iteration for σ_{ij}^2. In practice, we add a small regularization constant ϵ (e.g., 0.01) on the right-side of Eq. 7 to avoid the numerical problem of zero.

Another intuition is that the consensus between the estimate of $\mathbf{x}_i$ and $\mathbf{A}_{ij}\mathbf{x}_j + \mu_{ij}$ is equivalent to the consistency of the measurements $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$ and $\{\mathbf{z}_j, \mathbf{\Sigma}_j\}$. Therefore, when $\mathbf{z}_i$ and $\mathbf{z}_j$ are consistent, the estimate of $\mathbf{x}_i$ and $\mathbf{A}_{ij}\mathbf{x}_j + \mu_{ij}$ will be consensus, i.e., they will be almost the same. From Eq. 7, the estimate of σ_{ij}^2 will always approach to zero, i.e., zero is the only fixed-point. On the contrary, if they are inconsistent, then the estimate of $\mathbf{x}_i$ and $\mathbf{A}_{ij}\mathbf{x}_j + \mu_{ij}$ may deviate from each other, i.e., the convergent results of σ_{ij}^2 may be non-zero. This indicates that

there exists non-zero fixed-point for σ_{ij}^2. These motivate us for the following definition for inconsistency.

Definition 3.1 *If zero is the only fixed-point for σ_{ij}^2 in the Bayesian EM, $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$ and $\{\mathbf{z}_j, \mathbf{\Sigma}_j\}$ are* consistent*; if there exists non-zero fixed-points for σ_{ij}^2, they are* inconsistent*.*

This definition motivates us to detect the inconsistency by checking the convergent value of σ_{ij}^2. We thus have the following criterion to test consistency.

Criterion 3.2 *With a proper initialization, if the convergent results of σ_{ij}^2 in the Bayesian EM approaches to zero, then $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$ and $\{\mathbf{z}_j, \mathbf{\Sigma}_j\}$ are consistent. If it converges to a non-zero value, then they are inconsistent.*

In practice, a *proper* initialization should guarantee σ_{ij}^2 to converge to a non-zero fixed-point if there exists one, such a condition is necessary because zero is always a trivial fixed-point (see App. A). For better mathematical understanding of Definition 3.1, we proved the following Theorem 3.3 by studying the convergence of the Bayesian EM for pair-wise measurements. In Corollary 3.4, we also present a guidance to choose the *proper* initialization for Criterion 3.2 .

Theorem 3.3 *For a Markov network which models the integration of two sources, denote $\hat{\mathbf{z}}_2 = \mathbf{A}_{12}\mathbf{z}_2 + \mu_{12}$, $\hat{\mathbf{\Sigma}}_2 = \mathbf{A}_{12}\mathbf{\Sigma}_2\mathbf{A}_{12}^T$, $\mathbf{P} = \mathbf{\Sigma}_1 + \hat{\mathbf{\Sigma}}_2$ which is* real positive definite*, C_p the 2-norm conditional number and σ_{Pmax}^2 the largest eigenvalue of $\mathbf{P}$, and $\hat{\sigma}_{12}^2$ as the convergent results of σ_{12}^2 in the Bayesian EM. We have*

(a) *There exists a zero and at least one non-zero $\hat{\sigma}_{12}^2$ if*

$$\frac{1}{n}(\mathbf{z}_1 - \hat{\mathbf{z}}_2)^T\mathbf{P}^{-1}(\mathbf{z}_1 - \hat{\mathbf{z}}_2) \geq 2 + \sqrt{C_p} + \frac{1}{\sqrt{C_p}}. \quad (8)$$

(b) *$\hat{\sigma}_{12}^2$ can only be zero if*

$$\frac{1}{n}(\mathbf{z}_1 - \hat{\mathbf{z}}_2)^T\mathbf{P}^{-1}(\mathbf{z}_1 - \hat{\mathbf{z}}_2) < 4. \quad (9)$$

(c) *When there exists non-zero $\hat{\sigma}_{12}^2$, at least one of them is such that $0 < \hat{\sigma}_{12}^2 \leq \sigma_{Pmax}^2$*

The proof is presented in App. A. Highlighted by Theorem 3.3(c), we have the following corollary.

Corollary 3.4 *Under the same condition of Theorem 3.3, initializing σ_{12}^2 to be the largest eigen-value σ_{Pmax}^2 or the trace $T(\mathbf{P})$ of $\mathbf{P}$ in the Bayesian EM can guarantee a non-zero convergence for σ_{12}^2 if there exists one.*

The proof is presented in App. B. Theorem 3.3 and Corollary 3.4 provide a sound mathematical justification of Definition 3.1 about inconsistency and consistency. We denote the left side of Eq. 8 and Eq. 9 as $d(\mathbf{z}_1, \mathbf{z}_2)$, which is in fact a Mahalanobis distance. In principle, when $d(\mathbf{z}_1, \mathbf{z}_2)$ is too large, statistically $\{\mathbf{z}_1, \mathbf{\Sigma}_1\}$ and $\{\mathbf{z}_2, \mathbf{\Sigma}_2\}$ are significantly deviated from each other and thus they are inconsistent. In this case there exists at least one non-zero convergence of σ_{12}^2. On the other hand, if $d(\mathbf{z}_1, \mathbf{z}_2)$ is small, statistically $\{\mathbf{z}_1, \mathbf{\Sigma}_1\}$ and $\{\mathbf{z}_2, \mathbf{\Sigma}_2\}$ are not deviated from each other and thus they are consistent. Then there will be only zero convergence for σ_{12}^2.

Theorem 3.3(a) and (b) present two algebraic criteria (sufficient conditions) to judge if $\{\mathbf{z}_1, \mathbf{\Sigma}_1\}$ and $\{\mathbf{z}_2, \mathbf{\Sigma}_2\}$ are inconsistent or consistent, i.e., if Eq. 8 holds, then they are inconsistent, and they are consistent if Eq. 9 holds. The following remarks would make the understanding more clear:

- Since $B_c = 2 + \sqrt{C_p} + \frac{1}{\sqrt{C_p}} \geq 4$, if $4 \leq d(\mathbf{z}_1, \mathbf{z}_2) < B_c$, we can not directly tell if there exists a non-zero $\hat{\sigma}_{12}^2$. In other words, we can not immediately decide the consistency unless we run the Bayesian EM.
- In one dimensional case, i.e., $n = 1$, we have $B_c = 4$. Then the inconsistency/consistency of $\mathbf{z}_1$ and $\mathbf{z}_2$ can be determined by testing if $d(\mathbf{z}_1, \mathbf{z}_2) \gtreqless 4$.
- For $n \geq 2$, if C_p is good to be near 1, then B_c would be very close to 4. The interval $[4, B_c)$ would be very tight. Then either B_c or 4 can be approximately used for detecting inconsistency similar to the case $n = 1$.
- For $n \geq 2$, if C_p is not good to be very large, then $B_c \gg 4$. We must run the Bayesian EM with a proper initialization to judge the consistency when $d(\mathbf{z}_1, \mathbf{z}_2)$ falls in $[4, B_c)$.
- In a general setting, from Corollary 3.4, the largest eigenvalue or the trace of $\mathbf{\Sigma}_i + \mathbf{\Sigma}_j + \sum_{k \in \mathcal{N}(i,j)} \mathbf{\Sigma}_k$ is a *proper* initialization, where $\mathcal{N}(i, j)$ is the neighborhood of i and j. The trace is preferable since it can be more efficiently obtained.

4 Detection of inconsistency and falseness

Based on Criterion 3.2, let L_{ij} be the binary variable to indicate whether $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$ and $\{\mathbf{z}_j, \mathbf{\Sigma}_j\}$ are inconsistent, i.e., $L_{ij} = 1$ represents that they are and vice versa. Then the criterion to identify the inconsistency is

$$L_{ij} = \begin{cases} 0 & if\ \sigma_{ij}^2 \leq \epsilon \\ 1 & if\ \sigma_{ij}^2 > \epsilon \end{cases}, \quad (10)$$

where ϵ is the same regularization constant added in Eq. 7.

After the detection of inconsistency, the majority rule is adopted to determine if $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$ is false, i.e., if $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$ is inconsistent with the majority of its neighbors, then it is false, and vice versa. Without any other knowledge, the majority rule may be the best one to discriminate false measurements. The basic assumptions are that there are at least

three sources and the majority of the sources will obtain correct and thus consistent measurements. Since $\mathbf{O}_i$ is the binary variable to indicate if $\mathbf{z}_i$ is false, suppose part i has M_i neighbors, then

$$\mathbf{O}_i = \begin{cases} 0 & if \ \sum_{j\in\mathcal{N}(i)} L_{ij} \leq \lfloor \frac{M_i}{2} \rfloor \\ 1 & if \ \sum_{j\in\mathcal{N}(i)} L_{ij} > \lfloor \frac{M_i}{2} \rfloor \end{cases} \quad (11)$$

where $\lfloor \frac{M_i}{2} \rfloor$ is the largest integer that is not larger than $\frac{M_i}{2}$.

However, when the degrees (i.e., the number of neighboring nodes) of the nodes in the Markov network are highly unbalanced, the majority rule may fail even if there are less than 50% false measurements. One such example would be that the connections of $N > 6$ nodes form a circle and meanwhile the nodes $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are connected with all the other nodes. Then if the measurements $\mathbf{z}_1$, $\mathbf{z}_2$ and $\mathbf{z}_3$ are false and thus inconsistent with the others, all the other measurements will be regarded as "false" from Eq. 11.

Such a problem may not exist when the degrees of the nodes are well balanced. This reveals to us that in order to well exploit Eq. 11, we must construct a balanced Markov network to integrate the multiple sources, i.e., the degrees of the nodes must be close to one another.

5 Robust integration for visual tracking

Given all $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$s, we propose a two-stage robust integration approach:

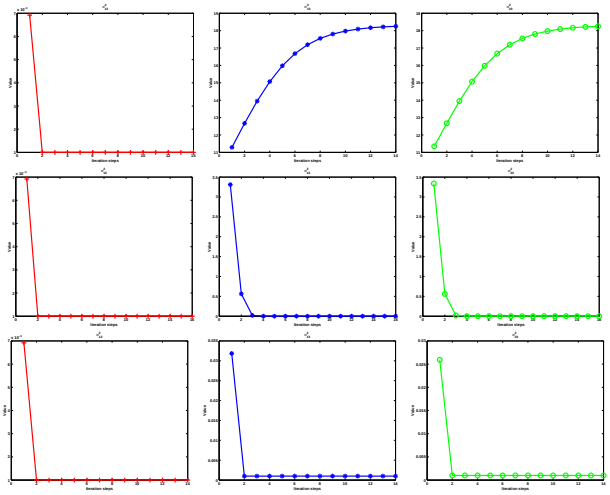
1. `False discrimination:` Perform the Bayesian EM on the original Markov network $\mathcal{M}_o$ defined by Eq. 1 and then identify the false measurements set $\mathbf{Z}_\mathcal{O}$ based on Eq. 11.

2. `Robust Integration:` Remove all $\mathbf{z}_i \in \mathbf{Z}_\mathcal{O}$, from $\mathcal{M}_o$. This forms the reduced Markov network $\mathcal{M}_r$ defined by Eq. 5. Perform the Bayesian EM on $\mathcal{M}_r$ to obtain the estimates for all $\mathbf{x}_i$ with $\mathbf{Z}_\mathcal{O}$ being removed from Eq. 6 and Eq. 7.

It is a completely *distributed* robust integration approach, where all the operations are performed individually at each node of the Markov network. After the false measurement at one source node i has been eliminated, as we can observe from Eq. 6 (eliminating $\mathbf{z}_i$ and $\mathbf{\Sigma}_i$ from it), the estimate of $\mathbf{x}_i$ will rely purely on the neighborhood estimates.

It can be immediately applied to part-based visual tracking, where $\psi(\mathbf{x}_i, \mathbf{x}_j)$ captures the structured constraints between two neighboring parts. It is also general to incorporate different tracking algorithms to obtain the part measurements $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}$, such as particle filtering [5] and flow based Lucas-Kanade tracker (LK) [9], etc..

There are three situations: (1) The measurements of all the parts are normal and consistent. (2) The measurements of some parts are *missing*, i.e., the $\phi(\mathbf{x}_i, \mathbf{z}_i)$ is a Gaussian with large co-variance. This might happen when the visual pattern of the target undergoes large variations but the visual model does not capture it well. (3) The measurements of some parts are inconsistent with those of the other. This implies that some measurements are false and it may be caused by either occlusion, clutter or camouflage in visual tracking. Our robust integration approach handles all these three situations in a unified way.



(a). σ_{12}^2 (b). σ_{13}^2 (c). σ_{23}^2

Figure 1. The change of σ_{ij}^2 in the Bayesian EM. First row: measurement $\mathbf{z}_3$ is false. Second row: measurement $\mathbf{z}_3$ is missing. Third row: all the measurements are consistent.

6 Experiments

6.1 Illustrative numerical example

We adopt a $2D$ numerical example to demonstrate how σ_{ij}^2 changes during the Bayesian EM. The Markov network models three sources, which are neighbors of one another. Without loss of generality, we set all $\mathcal{A}_{ij} = \mathbf{I}$ and $\mu_{ij} = 0$. In all the simulations, we fix $\mathbf{z}_1 = [2.1, 2.2]^T$, $\mathbf{z}_2 = [2.2, 2.1]^T$ and $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = [2.0, 1.0; 1.0, 2.0]$. We then set $\{\mathbf{z}_3, \mathbf{\Sigma}_3\}$ to be different values to simulate the three situations. Highlighted by Corollary 3.4, we always initialize all σ_{ij}^2 to be the trace of $\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2 + \mathbf{\Sigma}_3$.

We firstly simulate the case of false measurement, e.g., $\mathbf{z}_3 = [8.0, 9.0]^T$ and $\mathbf{\Sigma}_3 = [2.0, 1.0; 1.0, 2.0]$. It is obvious that $\{\mathbf{z}_3, \mathbf{\Sigma}_3\}$ is false. The changes of σ_{12}^2, σ_{13}^2 and σ_{23}^2 are presented in the first row of Fig. 1. As we can observe, σ_{12}^2 converges to 0.01, and both σ_{13}^2 and σ_{23}^2 converges to 18.25. Using Eq. 11, we easily identify $\mathbf{z}_3$ as

(a). #16 (b). #66 (c). #88 (d). #90 (e). #102 (f). #184

Figure 2. Rresults with flow measurement: the red, green, and yellow color denote false, normal and missing measurements, respectively.

a false measurement and it will be eliminated in the robust inference step. The MAP estimates before false elimination are $\mathbf{x}_1 = [2.83, 2.87]^T$, $\mathbf{x}_2 = [2.83, 2.87]^T$ and $\mathbf{x}_3 = [6.65, 7.55]^T$, which are erroneous and can be rectified after we eliminated $\mathbf{z}_3$.

We then simulate the case of missing measurement, e.g., $\mathbf{z}_3 = [8.0, 9.0]^T$ with $\mathbf{\Sigma}_3 = [10.0, 1.0; 1.0, 10.0]$. Although $\mathbf{z}_3$ is deviated from $\mathbf{z}_1$ and $\mathbf{z}_2$, its covariance $\mathbf{\Sigma}_3$ is pretty large so it is still consistent with the others. The changes of σ_{ij}^2 are presented in the second row of Fig. 1. We can observe that all of them converge to 0.01. In fact, the MAP estimates are $[2.89, 2.94]^T$ for all $\mathbf{x}_i$. We can see $\mathbf{z}_3$ has been counted far less than the other two measurements and the bias has largely been rectified in the estimates.

Last we simulate the easiest case where all the measurements are reliable and consistent, e.g., $\mathbf{z}_3 = [1.9, 1.8]^T$ with $\mathbf{\Sigma}_3 = [2.0, 1.0; 1.0, 2.0]$. The change of σ_{ij}^2 is presented in the third row of Fig. 1. Again, they all converge to 0.01 as expected. The final MAP estimates are $[2.07, 2.03]^T$ for all $\mathbf{x}_i$. We have extensively run the simulations with different settings. The results are coherent with what are presented.

6.2 Robust part based tracking

6.2.1 Part based tracking with LK tracker

(a). LK #90 (b). Holistic PF #108

Figure 3. Typical tracking failure (a). LK tracking frame #90. (b). Particle filtering with holistic appearance model frame #108

We first present the results using LK tracker [9] to obtain the part measurements. The test video clip is from the comedy cartoon "Tom and Jerry". The target is the poor cat Tom's face. Those "good features" [9] in Tom's face region are detected to be the node of the Markov network. The face region is manually cropped as a rectangle in the first frame.

Each node is associated with a 7×7 image patch (the appearance model) centering at the feature point, and it is connected with the three nearest nodes. The $\mathbf{x}_i$ is the $2D$ position of the ith good feature. At the current frame, we set $\mathbf{A}_{ij} = \mathbf{I}_2$ and set μ_{ij} to be the relative position of part i and j in the previous frame. Each $\mathbf{z}_i$ is obtained by the flow based LK tracker. The $\mathbf{\Sigma}_i$ is obtained by evaluating the response distribution using SSD similar to that in [15].

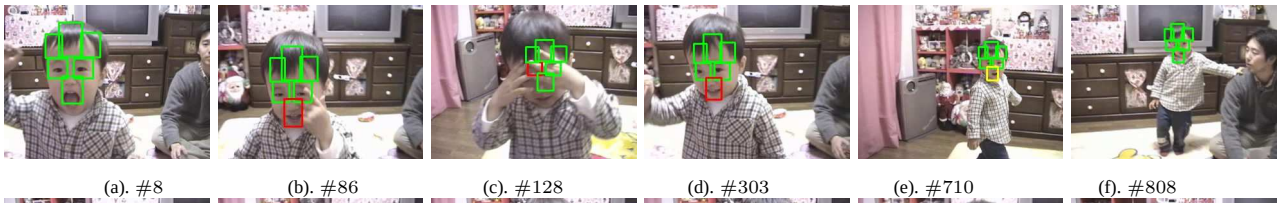
We show some sample results in Fig. 2 (detailed results in "330.wmv"). Our algorithm successfully identifies the false, missing and normal measurements, as shown in red, yellow and green, respectively. The video has 187 frames and our algorithm obtains robust results. With 50 parts, it runs at 10 frames/second without code optimization.

The pure LK tracker and the particle filtering (PF) with a holistic appearance model are easy to fail in this video clip. We show the typical failure cases in Fig. 3. The failures are due to the dramatic expression change (Fig. 3(a)), the sudden view changes and abrupt motion of Tom (Fig. 3(b)). The number of particles for holistic PF is 200 and all algorithms are initialized with the same rectangle.

6.2.2 Part based tracking with particle filtering

In this section, we present the tracking results using particle filtering [5] to obtain the part measurement. The $\mathbf{x}_i$ is four dimensional (two for translations and two for scalings). The target parts are selected manually and a fully connected Markov network is adopted. The $\mathbf{A}_{ij}$ and μ_{ij} are estimated from some manually annotated images by least square fitting. There is a residue error σ_{ij0}^2 from the least square fitting. It was used as the σ_{ij}^2 in the robust integration step, i.e., after removing the false measurements, we fix $\sigma_{ij}^2 = \sigma_{ij0}^2$ and perform the Bayesian inference using Eq. 6. Note each component has a template image patch to build the appearance based likelihood model $\phi(\mathbf{x}_i, \mathbf{z}_i)$. The mean estimates and the covariances of the posterior particle sets are adopted as the part measurements $\{\mathbf{z}_i, \mathbf{\Sigma}_i\}s$.

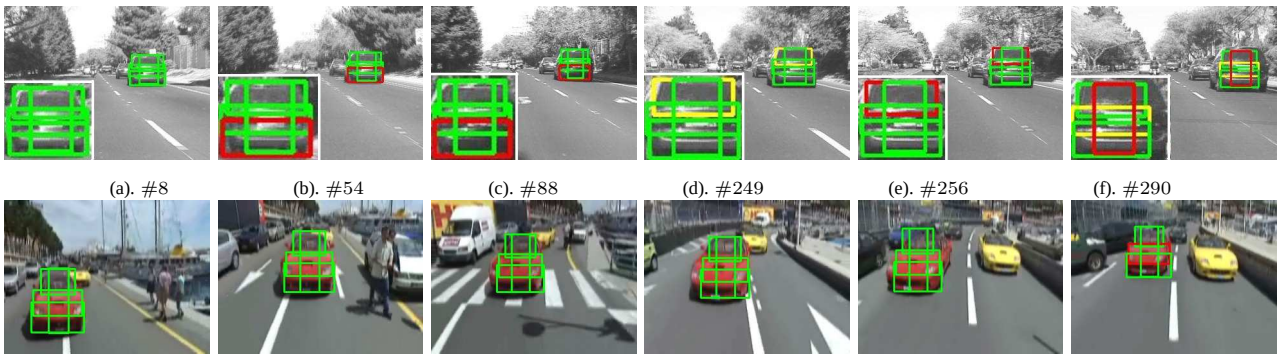
We present sample results on different video sequences in Fig. 4 and Fig. 5. These test videos are typical, where



(a). #8 (b). #86 (c). #128 (d). #303 (e). #710 (f). #808

(a). #116 (b). #128 (c). #138 (d). #144 (e). #148 (f). #154

Figure 4. Comparison of robust integration by the proposed approach and blind integration without inconsistency detection and false elimination – First row: Proposed integrating approach (green-normal, red-false, yellow-missing). Second row: Blind integrating.



(a). #8 (b). #54 (c). #88 (d). #249 (e). #256 (f). #290

(a). #280 (b). #306 (c). #310 (d). #316 (e). #318 (f). #399

Figure 5. Results with PF measurement: Results in the first row are enlarged for better visual quality (green-normal, red-false, yellow-missing).

the targets present large appearance variations due to the significant view, scale, lighting changes and the presence of occlusions. Fig. 4 shows the results of tracking the face of a kid. The first row of Fig. 4 shows the results of the proposed approach, where inconsistent measurements are detected and those false ones are eliminated. For comparison, the second row of Fig. 4 shows the results of blind integration without inconsistency detection and false elimination. Note how the tracking results have been distracted due to the integration of those false measurements during occlusion. The video has 820 frames.

In Fig. 5, we present the results on two car video sequences, which have 348 and 399 frames, respectively. Detailed video results are presented in "330.wmv". We also tested the accuracy of the results shown in the first row of Fig. 5 on the two translation parameters. 300 frames are labeled and the centroid points of the labeled rectangle is adopted as the ground truth. For the tracking results, the

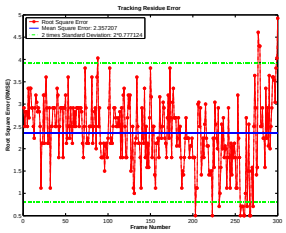
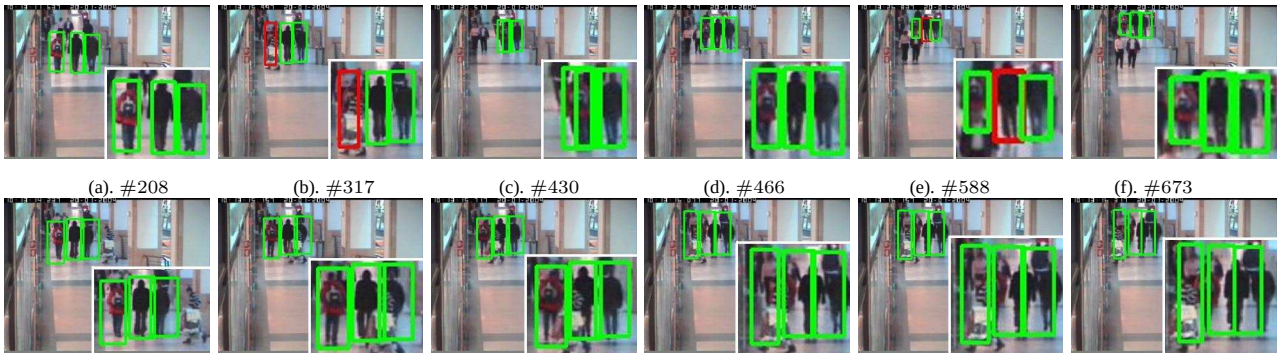


Figure 6. Root square error of the results on the car sequences in the first row of Fig. 5.

centroid point of all the part rectangles is used as the overall translation parameters. We then calculate the root square error at each frame, as shown in Fig. 6. The root mean square error is 2.36 pixels with stand deviation 0.78 on 320×240 images. This shows the accuracy of the proposed approach.



(a). #208 (b). #317 (c). #430 (d). #466 (e). #588 (f). #673

(a). #280 (b). #306 (c). #310 (d). #316 (e). #318 (f). #322

Figure 7. Tracking a group of persons: First row: Our integrating approach (green-normal, red-false, yellow-missing). Second row: Blind integration. The results are enlarged for better visual quality.

6.2.3 Tracking a group of objects

A direct generalization of part based tracking is to track several objects moving in a group. We tested the robust integration on part of a video sequence[1] where three persons walking in a corridor of a shopping mall. Again, a fully connected Markov network is adopted and the measurement of each person is obtained by particle filter. Some sample results are presented in the first row of Fig. 7. In Fig. 7(b) (first row), frame #317, the left person has been occluded by another person and the measurement is false. Our algorithm clearly identified and corrected it. For comparison, we also present the results by blind integration in the second row of Fig. 7. Note how it fails due to occlusion. The video sequence of the three persons has 697 frames. More video results can be found in "330.wmv".

7 Conclusions and future work

We proposed a novel distributed framework for detecting and integrating of inconsistent measurements. The modeling is based on Markov networks. The Bayesian EM inference reveals the iterative integration of the measurements, from which principled criteria were developed to detect inconsistency. We regard measurements which are inconsistent with the majority of their neighbors as false. They will be eliminated and the integration is performed again, i.e., the estimates in those nodes with false measurements will only rely on the measurements from its neighbors. We applied the proposed robust integration framework for part based visual tracking and promising results were obtained.

Future work may include the automatic part selection, and better means to handle the integration in unbalanced Markov networks. We are also interested in exploiting the integration framework to other vision applications.

[1]From the EC Funded CAVIAR project/IST 2001 37540, found at URL: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/

Acknowledgments

This work was supported in part by NSF IIS-0347877, IIS-0308222, and Murphy and Richter Fellowships for GH.

A Proof of Theorem 3.3

Proof Fixing σ_{12}^2, Eq. 6 guarantees to iteratively obtain the exact MAP estimate on the joint posterior Gaussian. We denote $\hat{\mathbf{x}}_2 = \mathbf{A}_{12}\mathbf{x}_2 + \mu_{12}$ and $\mathbf{S} = \mathbf{P} + \sigma_{12}^2\mathbf{I}$. The convergent results in the E-Step in Eq. 6 is the same as,

$$\begin{bmatrix} \mathbf{x}_1 \\ \hat{\mathbf{x}}_2 \end{bmatrix} = \begin{bmatrix} (\sigma_{12}^2\mathbf{I} + \hat{\mathbf{\Sigma}}_2)\mathbf{S}^{-1}\mathbf{z}_1 + \mathbf{\Sigma}_1\mathbf{S}^{-1}\hat{\mathbf{z}}_2 \\ \hat{\mathbf{\Sigma}}_2\mathbf{S}^{-1}\mathbf{z}_1 + (\sigma_{12}^2\mathbf{I} + \mathbf{\Sigma}_1)\mathbf{S}^{-1}\hat{\mathbf{z}}_2 \end{bmatrix}. \quad (12)$$

Embedding it to the M-Step in Eq. 7, we have

$$\sigma_{12}^2 = \frac{1}{n}\sigma_{12}^2\sigma_{12}^2(\mathbf{z}_1 - \hat{\mathbf{z}}_2)^T\mathbf{S}^{-1}\mathbf{S}^{-1}(\mathbf{z}_1 - \hat{\mathbf{z}}_2). \quad (13)$$

Since zero is a solution of σ_{12}^2 for Eq. 13, we only need to analyze the existence of non-zero solutions of σ_{12}^2 for

$$\frac{1}{n}\sigma_{12}^2(\mathbf{z}_1 - \hat{\mathbf{z}}_2)^T\mathbf{S}^{-1}\mathbf{S}^{-1}(\mathbf{z}_1 - \hat{\mathbf{z}}_2) - 1 = 0. \quad (14)$$

Since $\mathbf{P}$ is *real positive definite*, there exists an orthonormal matrix $\mathbf{Q}$ such that $\mathbf{P} = \mathbf{Q}\mathbf{D}_p\mathbf{Q}^T$ where $\mathbf{D}_p = diag[\sigma_1^2, \ldots, \sigma_n^2]$ and $\sigma_1^2 \geq \ldots \geq \sigma_n^2 > 0$. Let $C_p = \frac{\sigma_1^2}{\sigma_n^2}$. We then have $\mathbf{S} = \mathbf{Q}\mathbf{D}_s\mathbf{Q}^T$ and $\mathbf{S}^{-1} = \mathbf{Q}^T\mathbf{D}_s^{-1}\mathbf{Q}$, where $\mathbf{D}_s = diag[\sigma_1^2 + \sigma_{12}^2, \ldots, \sigma_n^2 + \sigma_{12}^2]$ and $\mathbf{D}_s^{-1} = diag[\frac{1}{\sigma_1^2+\sigma_{12}^2}, \ldots, \frac{1}{\sigma_n^2+\sigma_{12}^2}]$. Denote $\tilde{\mathbf{z}} = \mathbf{Q}(\mathbf{z}_1 - \hat{\mathbf{z}}_2) = [\tilde{z}_1, \ldots, \tilde{z}_n]^T$, we have

$$\frac{1}{n}\sigma_{12}^2(\mathbf{z}_1 - \hat{\mathbf{z}}_2)^T\mathbf{S}^{-2}(\mathbf{z}_1 - \hat{\mathbf{z}}_2) = \frac{1}{n}\sum_{i=1}^{n}\frac{\sigma_{12}^2\tilde{z}_i^2}{(\sigma_i^2 + \sigma_{12}^2)^2} \quad (15)$$

$$\frac{1}{n}(\mathbf{z}_1 - \hat{\mathbf{z}}_2)^T \mathbf{P}^{-1}(\mathbf{z}_1 - \hat{\mathbf{z}}_2) = \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{z}_i^2}{\sigma_i^2}. \quad (16)$$

From Eq 15, we only need to analyze the solution of σ_{12}^2 for

$$F(\sigma_{12}^2) = \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{z}_i^2}{\sigma_i^2} \cdot \frac{1}{2 + \frac{\sigma_i^2}{\sigma_{12}^2} + \frac{\sigma_{12}^2}{\sigma_i^2}} - 1 = 0. \quad (17)$$

We proceed to prove the three cases in Theorem 3.3.

(a). Eq. 8 means $d = \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{z}_i^2}{\sigma_i^2} > 2 + \sqrt{\frac{\sigma_1^2}{\sigma_n^2}} + \sqrt{\frac{\sigma_n^2}{\sigma_1^2}} \geq 4$. When $\sigma_{12}^2 = k_1 = (d-2)\sigma_1^2$, for any i, we have $\frac{1}{2+\frac{\sigma_i^2}{\sigma_{12}^2}+\frac{\sigma_{12}^2}{\sigma_i^2}} < \frac{1}{2+0+d-2} = \frac{1}{d}$. Thus $F(k_1) < \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{z}_i^2}{\sigma_i^2} \cdot \frac{1}{d} - 1 = 0$. When $\sigma_{12}^2 = k_2 = \sqrt{\sigma_1^2 \sigma_n^2}$, for any i, $\frac{1}{2+\frac{\sigma_i^2}{\sigma_{12}^2}+\frac{\sigma_{12}^2}{\sigma_i^2}} \geq \frac{1}{2+\frac{\sigma_n^2}{k_2}+\frac{k_2}{\sigma_1^2}} = \frac{1}{2+\sqrt{\frac{\sigma_1^2}{\sigma_n^2}}+\sqrt{\frac{\sigma_n^2}{\sigma_1^2}}} \geq \frac{1}{d}$, thus $F(k_2) \geq \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{z}_i^2}{\sigma_i^2} \cdot \frac{1}{d} - 1 = 0$. Since $0 < k_2 < k_1$ and $F(\cdot)$ is continuous, there exists a k_3 where $k_2 \leq k_3 < k_1$ and $F(k_3) = 0$. This proves Theorem 3.3(a).

(b). Eq. 9 means $d = \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{z}_i^2}{\sigma_i^2} < 4$, then $F(\sigma_{12}^2) \leq \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{z}_i^2}{\sigma_i^2} \cdot \frac{1}{4} - 1 = \frac{d}{4} - 1 < 0$ for all $\sigma_{12}^2 > 0$. Thus Eq. 17 has no non-zero solution. Theorem 3.3(b) is proven.

(c). Let $F(\sigma_M^2) = \max F(\sigma_{12}^2)$, we show that it must be such that $\sigma_n^2 \leq \sigma_M^2 \leq \sigma_1^2$. Define $F_i(\sigma_{12}^2) = \frac{1}{n}\frac{\tilde{z}_i^2}{\sigma_i^2}\frac{1}{2+\frac{\sigma_i^2}{\sigma_{12}^2}+\frac{\sigma_{12}^2}{\sigma_i^2}}$ thus $F(\sigma_{12}^2) = \sum_i F_i(\sigma_{12}^2) - 1$. Each $F_i(\sigma_{12}^2)$ is monotonically increasing for $0 < \sigma_{12}^2 \leq \sigma_i^2$ and monotonically decreasing for $\sigma_{12}^2 \geq \sigma_i^2$. Therefore $F(\sigma_{12}^2)$ must be monotonically increasing for $0 < \sigma_{12}^2 \leq \sigma_n^2$ and monotonically decreasing for $\sigma_{12}^2 \geq \sigma_1^2$. This tells us that the global maximum of $F(\sigma_{12}^2)$ can only be taken in $\sigma_n^2 \leq \sigma_{12}^2 \leq \sigma_1^2$, thus $\sigma_n^2 \leq \sigma_M^2 \leq \sigma_1^2$.

The existence of a non-zero convergent value of σ_{12}^2 implies a non-zero solution for Eq. 17. We have $F(\sigma_M^2) \geq 0$ otherwise $F(\sigma_{12}^2) < 0$ for all σ_{12}^2 and there is no solution for Eq. 17. Since $F(0) \to -1$ and $F(\sigma_{12}^2)$ is continuous, there must exist a k_4 such that $0 < k_4 \leq \sigma_M^2 \leq \sigma_1^2$ and $F(k_4) = 0$. This immediately proves Theorem 3.3(c). ∎

B Proof of Corollary 3.4

Proof The Bayesian EM constitutes a fixed-point iteration of σ_{12}^2 in Eq. 13. From Theorem 3.3(c), when non-zero fixed-points exist, at least one of them, $\hat{\sigma}_{12}^2$, is such that $0 < \hat{\sigma}_{12}^2 \leq \sigma_{Pmax}^2 < T(\mathbf{P})$. Then, if the fixed-point iteration is initialized at σ_{pmax}^2 or $T(\mathbf{P})$, it can never surpass $\hat{\sigma}_{12}^2$ to converge to zero since they are scalars. This indicates that σ_{pmax}^2 and $T(p)$ are proper initialization for σ_{12}^2. ∎

References

[1] D. Comaniciu. Nonparametric information fusion for motion estimation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2003.

[2] W. T. Freeman and E. C. Pasztor. Learning low-level vision. In *Proc. IEEE International Conference on Computer Vision*, pages 1182–1189, 1999.

[3] B. Heisele, T. Serre, M. Pontil, and T. Poggio. Component-based face detection. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 657–662, 2001.

[4] S. Ioffe and D. A. Forsyth. Mixtures of trees for object recognition. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 180–185, 2001.

[5] M. Isard and A. Blake. Contour tracking by stochastic propagation of conditional density. In *Proc. European Conference on Computer Vision*, volume 1, pages 343–356, 1996.

[6] S. Julier and J. Uhlmann. A nondivergent estimation algorithm in the presence of unknown correlations. In *Proc. of the American Control Conference*, Alberqueque, 6 1997.

[7] X. R. Li, Y. Zhu, J. Wang, and C. Han. Optimal linear estimation fusionłpart i: Unified fusion rules. *IEEE Transaction on Information Theory*, 49(9):2192–2208, 2003.

[8] V. I. Pavlovic. *Dynamic Bayesian Networks for Information Fusion with Application to Human-Computer Interfaces*. Phd thesis, Department of Electrical and Computer Engieering, University of Illinois at Urbana-Champaign, 1999.

[9] J. Shi and C. Tomasi. Good features to track. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 593–600, June 1994.

[10] L. Sigal, S. Bhatia, S. Roth, and M. Black. Tracking loose-limbed people,. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 421–428, 2004.

[11] K. Toyama and E. Horvitz. Bayesian modality fusion: Probabilistic integration of multiple vision algorithms for head tracking. In *Proc. of Fourth Asian Conference on Computer Vision*, Taiwan, January 2000.

[12] J. K. Uhlmann. Covariance consistency methods for fault-tolerant distributed data fusion. *Information Fusion, Elsevier Science*, 4(3):201–215, 3 2003.

[13] Y. Wu, G. Hua, and T. Yu. Tracking articulated body by dynamic markov network. In *Proc. IEEE International Conference on Computer Vision*, pages 1094–1101, Nice,Côte d'Azur,France, October 2003.

[14] Y. Wu and T. S. Huang. Robust visual tracking by co-inference learning. In *Proc. IEEE Int'l Conference on Computer Vision*, volume II, pages 26–33, 2001.

[15] X. S. Zhou, D. Comaniciu, and A. Gupta. An information fusion framework for robust shape tracking. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 27(1):115–129, 2005.