Game-Theoretic Multiple Target Tracking

Ming Yang[†], Ting Yu[‡], Ying Wu[†]
[†]Dept. of EECS, Northwestern Univ.
2145 Sheridan Rd., Evanston, IL 60208
mya671,yingwu@ece.northwestern.edu

[‡]VCV Lab, GE Global Research
One Research Circle, Niskayuna, NY 12309
yut@ge.com

Abstract

Video-based multiple target tracking (MTT) is a challenging task when similar targets are present in close vicinity. Because their visual observations are mixed and difficult to segment, their motions have to be estimated jointly. Most existing approaches perform this joint motion estimation in a centralized fashion and involve searching a rather high dimensional space, and thus leading to quite complicated joint trackers. This paper brings a new view to MTT from a game-theoretic perspective, bridging the joint motion estimation and the Nash Equilibrium of a game. *Instead of designing a centralized tracker, MTT is decentralized and a set of individual trackers is used, each of which tries to maximize its visual evidence for explaining its motion as well as generates interferences to others. Modelling this competition behavior, a special* game *is designed so that the difficult joint motion estimation is achieved at the Nash Equilibrium of this game where no individual tracker has incentives to change its motion estimate. This paper substantializes this novel idea in a solid case study where individual trackers are kernel-based trackers. An efficient best response updating procedure is designed to find the Nash Equilibrium. The powerfulness of this game-theoretic MTT is shown by promising results on difficult real videos.*

1. Introduction

Multiple target tracking (MTT) in video is a critical and fundamental task in many real applications, *e.g.* video surveillance, vision-based interfaces, and video analysis. This task would not have been more difficult than tracking a single target, if multiple targets had quite different visual appearances or were not present in close vicinity. In practice, however, it is very common that those targets may look similar and may occlude each other in video during their interactions. As a result, it is understandable that losing tracks and associating wrong tracks to some targets are common experiences of the failures in vision-based MTT systems.

The challenge roots in the difficulty that estimating the motions of multiple targets cannot be treated independently if they are present in close vicinity, because their visual observations (or visual evidence) are mixed and it is generally very difficult, if not impossible, to figure out the right associations of these observations to the individuals targets (that implies a general segmentation problem). To handle this difficulty, the motions of multiple targets have to be jointly estimated from the mixed visual observations, which makes MTT much more difficult than tracking a single target as the solution space of MTT is much larger.

This joint estimation problem can be performed in a centralized fashion by formulating a joint observation model, as treated in many existing methods [13, 9, 12, 7, 6, 11, 14, 19, 8]. Because the joint observation model evaluates hypotheses of joint motion states, these methods lead to complicated centralized MTT trackers that generally need to search a rather high dimensional solution space.

This paper brings a new view to MTT from a game-theoretic perspective, bridging the joint motion estimation and the Nash Equilibrium of a game. It advocates a decentralized methodology that solves MTT through the competition among a set of simple individual target trackers. These individual trackers compete against each other for visual observations, and each individual tracker tries to maximize its visual evidence for explaining its motion and also generates interferences to other individual trackers. This can naturally be formulated as a *game* in which individual trackers are *players*, each of which estimates its own motion (*i.e.*, choosing its own *strategy*) by optimizing its own objective (*i.e.*, *utility* or *payoff*). The solution to MTT is tied to the *Nash Equilibrium* (N.E.) [10] of the game, where no player can achieve a better payoff by choosing a different strategy.

The objective functions for the individual trackers cannot be arbitrarily chosen, for example based on intuitions or heuristics, as they characterize the game and its Nash Equilibrium and thus influencing the solution to MTT. To make this clear, specifically, this paper presents a solid and novel case study where individual trackers are kernel-based trackers [2, 3]. Based on the kernel representation, we introduce

an *interference model* that describes the visual observations of the individual tracker by considering the interferences generated from other trackers, and then define a joint motion estimation problem. The Karush-Kuhn-Tucker (KKT) conditions of this joint optimization produce a fixed-point equation. Naive iteration is not likely to reach the fixed-point, as it may not converge. Therefore, inspired by the *supermodular game* theory, we construct a game whose Nash Equilibrium corresponds to the fixed-point of the KKT conditions. More important, we design an efficient iterative best-response updating procedure that guarantees the convergence to the N.E. under certain conditions and this is provable. This best-response updating is done in a closed form thus it is quite computationally appealing.

The proposed game-theoretic MTT method has many merits. First, it is decentralized as each individual tracker only needs to optimize its own objective, and the complicated joint motion estimation is avoided. This decentralized scheme greatly reduces the computational complexity. In addition, the individual motion estimation is computed in a simple closed form and is computationally very efficient. Moreover, the proposed method is theoretically plausible because of its convergence properties.

2. Related work

Multiple target tracking has been studied extensively in literature and can be back-traced to [13]. Most work assume that one target hypothesis can only claim a single image observation and one observation can only support one hypothesis. This assumption can be referred as a probabilistic exclusion principle [9] and used as a prior in the well-known joint probabilistic data association filter (JPDAF) [1, 12] and multiple hypothesis tracking (MHT) [4]. Thus, the key problem in multiple target tracking is to infer the optimal joint motion configuration in a high dimensional space. This can be done in a centralized fashion by sampling or sequential Monte Carlo [9, 14, 7, 6, 19, 11, 8], or evolutionary optimization [5], or in a decentralized manner [18] by inferring on a Markov Network. Object detectors may also be included [11, 17].

Different from these existing methods, we bridge the joint motion estimation and the Nash Equilibrium of a game. We construct a *non-cooperative game* [10, 15] that characterizes the competition among a set of individual trackers. The Nash Equilibrium of this game corresponds to a local optimum of the joint motion configuration and can be solved by an efficient decentralized method.

3. Interference model

In this section, we introduce a new analytical interference model for kernel-based trackers, which is a key component in formulating the game-theoretic MTT. This interference model takes both target appearances and spatial relations into consideration.

3.1. Joint likelihood maximization

Denote the motion parameters for the ith target by θ_i. Its corresponding support is denoted by Ω_i, *i.e.* the set of pixels $\{\mathbf{x}_n\}$ within the region of target i. Thus, the motions of a number of N targets can be estimated by maximizing the joint likelihood,

$$\Theta^* = \underset{\{\theta_1,\cdots,\theta_N\}}{\operatorname{argmax}} P(\bigcup_{i=1}^{N} \Omega_i|\theta_1,\cdots,\theta_N). \quad (1)$$

If no occlusion is present, *i.e.* $\Omega_i \cap \Omega_j = \emptyset, \forall i,j \leq N$. This joint optimization can be done independently:

$$\theta_i^* = \underset{\theta_i}{\operatorname{argmax}} P(\Omega_i|\theta_i), \quad \forall i \leq N. \quad (2)$$

If occlusion is present, *i.e.* $\Omega_i \cap \Omega_j \neq \emptyset, \exists i,j \leq N$, we can assign the pixels in the overlapped regions to different targets probabilistically, thus

$$\Theta^* = \underset{\{\theta_1,\cdots,\theta_N\}}{\operatorname{argmax}} \prod_{i=1}^{N} P(\hat{\Omega}_i|\theta_1,\cdots,\theta_N), \quad (3)$$

where $\hat{\Omega}_i$ is the probabilistic support of target i. This is equivalent to an energy minimization problem:

$$\Theta^* = \underset{\{\theta_1,\cdots,\theta_N\}}{\operatorname{argmin}} -\sum_{i=1}^{N} \ln P(\hat{\Omega}_i|\theta_1,\cdots,\theta_N). \quad (4)$$

3.2. Kernel-based likelihood

Specifically, for a kernel-based tracker, a target is represented by a kernel weighted feature histogram [2]. The motion parameters are denoted by $\theta \triangleq \{\mathbf{y}, h\}$, where $\mathbf{y}$ is the location of the kernel center and h is its scale. Denote by $\mathbf{x}_n$ the 2D pixel location and $z_n \triangleq ||\frac{\mathbf{x}_n - \mathbf{y}}{h}||$. The kernel function $k(z_n^2)$ in this paper is the Epanechnikov kernel:

$$k(z_n^2) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-z_n^2), & z_n^2 < 1 \\ 0, & \text{otherwise} \end{cases}, \quad (5)$$

where $d = 2$ and c_d is the area of the unit circle. The negative derivative of the kernel is denoted by $g(z_n^2) \triangleq -k'(z_n^2)$.

Following the notations in [2], for a single tracker without interference, the model of target i is described by an M-bin histogram $\mathbf{q}_i = \{q_{im}\}_{m=1,\cdots,M}$, and the target hypothesis by $\mathbf{p}_i(\mathbf{y}_i) = \{p_{im}(\mathbf{y}_i)\}_{m=1,\cdots,M}$,

$$p_{im}(\mathbf{y}_i) = \sum_{\mathbf{x}_n \in \Omega_i} k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)\delta[b(\mathbf{x}_n) - m], \quad (6)$$

where $\delta[\cdot]$ is the Kronecker delta function and the function $b(\cdot)$ maps the pixel location $\mathbf{x}_n$ to a bin index m. The Bhattacharyya coefficient $\rho(\mathbf{y}_i)$ is employed to measure the similarity between a target hypothesis and the model

$$\rho(\mathbf{y}_i) = \sum_{m=1}^{M} \sqrt{p_{im}(\mathbf{y}_i) q_{im}}. \tag{7}$$

Since the distance from the hypothesis histogram $\mathbf{p}_i(\mathbf{y}_i)$ to the model histogram $\mathbf{q}_i$ can be defined as $d(\mathbf{y}_i) = \sqrt{1-\rho(\mathbf{y}_i)}$, the likelihood model for tracker i (in Eq. 2) without considering interference can be formulated as:

$$P(\Omega_i|\theta_i) \propto e^{1-\rho(\mathbf{y}_i)}. \tag{8}$$

3.3. Kernel-based interference model

Due to partial occlusion, we need to consider the interference among the N targets, *i.e.* $\Omega_i \cap \Omega_j \neq \emptyset, \exists i,j \leq N$. The observation model for tracker i is no longer solely determined by $\mathbf{y}_i$ but the joint motion configuration of all trackers (which is denoted by $\{\mathbf{y}_i, \mathbf{y}_{-i}\} = \{\mathbf{y}_i, \cdots, \mathbf{y}_N\}$ to highlight other trackers' interference to tracker i). In view of this, we generalize the kernel-based histogram model by,

$$\hat{p}_{im}(\mathbf{y}_i, \mathbf{y}_{-i}) = \frac{1}{C_i} \sum_{\mathbf{x}_n \in \Omega_i} \left\{ k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2) \delta[b(\mathbf{x}_n) - m] \cdot \frac{q_{im}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)}{\sum_{j=1}^{N} q_{jm}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_j}{h_j}||^2)} \right\}, \tag{9}$$

where $C_i \leq 1$ is a normalization term. The probability that the pixel $\mathbf{x}_n$ is within Ω_i is approximated by

$$\frac{q_{im}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)}{\sum_{j=1}^{N} q_{jm}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_j}{h_j}||^2)}, \tag{10}$$

where $q_{im}(\mathbf{x}_n) = \sum_{m=1}^{M} q_{im}[\delta(b(\mathbf{x}_n) - m)]$ is the histogram bin value for pixel $\mathbf{x}_n$ in the target model $\mathbf{q}_i$. Please note when using Epanechnikov kernel with a finite support, if one tracker has no overlap with others, Eq. 9 degenerates to Eq. 6. To avoid numerical problems, we set $q_{im} = \epsilon > 0, \forall m < M$, where ϵ is a very small value, to guarantee non-zero bins $q_{im}(\mathbf{x}_n)$ and $q_{jm}(\mathbf{x}_n)$.

The *generalized Bhattacharyya coefficient* is defined as $\hat{\rho}(\mathbf{y}_i, \mathbf{y}_{-i}) = \sum_{m=1}^{M} \sqrt{\hat{p}_{im}(\mathbf{y}_i, \mathbf{y}_{-i}) q_{im}}$. Then, the likelihood model for target i with interference is formulated as:

$$P(\hat{\Omega}_i|\theta_1, \cdots, \theta_N) \propto e^{1-\hat{\rho}(\mathbf{y}_i, \mathbf{y}_{-i})}. \tag{11}$$

This interference model takes both the appearance similarity and spatial relations into account. This interference model down-weights those pixels that are in the overlapped regions of different trackers and have ambiguous identities.

4. Game-theoretic multiple target tracking

Based on the interference model, we can formulate the joint motion estimation (Sec. 4.1) and construct a game (Sec. 4.2) whose N.E. corresponds to a local optimum of the joint motion estimation and can be efficiently solved (Sec. 4.3). The algorithm is summarized in Sec. 4.4.

4.1. Joint motion estimation

Assuming that the scales remain constant when multiple targets approach to each other, based on the interference likelihood model (Eq. 11), the minimization of the joint energy (in Eq. 4) is equivalent to:

$$\max_{\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}} J_1(\mathbf{y}_1, \cdots, \mathbf{y}_N) = \sum_{i=1}^{N} \hat{\rho}_i(\mathbf{y}_i, \mathbf{y}_{-i}). \tag{12}$$

Maximizing the joint likelihood is equivalent to optimizing the joint kernel locations of all targets that maximize the sum of the generalized Bhattacharyya coefficients.

Denote the initial locations of the trackers by $\{\mathbf{y}_i^0, \mathbf{y}_{-i}^0\}$. Then, performing Taylor expansion *w.r.t.* $\hat{p}_{im}(\mathbf{y}_i^0, \mathbf{y}_{-i}^0)$ and plugging Eq. 9 into $\hat{\rho}_i(\mathbf{y}_i, \mathbf{y}_{-i})$, $\hat{\rho}_i(\mathbf{y}_i, \mathbf{y}_{-i})$ can be approximated by

$$\begin{aligned}
&\hat{\rho}_i(\mathbf{y}_i, \mathbf{y}_{-i}) = \sum_{m=1}^{M} \sqrt{\hat{p}_{im}(\mathbf{y}_i, \mathbf{y}_{-i}) q_{im}} \\
&\approx \frac{1}{2} \sum_{m=1}^{M} \left(\sqrt{\hat{p}_{im}(\mathbf{y}_i^0, \mathbf{y}_{-i}^0) q_{im}} + \hat{p}_{im}(\mathbf{y}_i, \mathbf{y}_{-i}) \sqrt{\frac{q_{im}}{\hat{p}_{im}(\mathbf{y}_i^0, \mathbf{y}_{-i}^0)}} \right) \\
&= \frac{1}{2} \sum_{m=1}^{M} \sqrt{\hat{p}_{im}(\mathbf{y}_i^0, \mathbf{y}_{-i}^0) q_{im}} + \\
&\frac{1}{2C_i} \sum_{\Omega_i} \omega_i(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2) \frac{q_{im}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)}{\sum_{j=1}^{N} q_{jm}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_j}{h_j}||^2)},
\end{aligned} \tag{13}$$

where $\omega_i(\mathbf{x}_n)$ is determined by the initial status of tracker i $\hat{p}_{im}(\mathbf{y}_i^0, \mathbf{y}_{-i}^0)$ and the model histogram $\mathbf{q}_i$ of target i,

$$\omega_i(\mathbf{x}_n) = \sum_{m=1}^{M} \delta[b(\mathbf{x}_n) - m] \sqrt{\frac{q_{im}}{\hat{p}_{im}(\mathbf{y}_i^0, \mathbf{y}_{-i}^0)}}. \tag{14}$$

Since only the second term in Eq. 13 is related to the variable $\{\mathbf{y}_i, \mathbf{y}_{-i}\}$ given the initial locations, we can ignore the terms in J_1 that are not affected by $\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}$. Then we redefine the objective function and have:

$$\max_{\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}} J_2(\mathbf{y}_1, \cdots, \mathbf{y}_N) \triangleq \sum_{i=1}^{N} r_i(\mathbf{y}_i, \mathbf{y}_{-i}), \tag{15}$$

where $r_i(\mathbf{y}_i, \mathbf{y}_{-i})$ corresponds to the individual matching

of tracker i (as the second term in Eq. 13):

$$r_i(\mathbf{y}_i, \mathbf{y}_{-i}) \triangleq \frac{1}{2C_i} \sum_{\Omega_i} \frac{\omega_i(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)}{1 + \sum_{j=1, j\neq i}^{N} \frac{q_{jm}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_j}{h_j}||^2)}{q_{im}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)}}. \quad (16)$$

Since ∇J_2 *w.r.t.* to $\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}$ is intractable, we further approximate it with a lower bound $J_3 \leq J_2$:

$$\max_{\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}} J_3(\mathbf{y}_1, \cdots, \mathbf{y}_N) \triangleq \sum_{i=1}^{N} \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i}), \quad (17)$$

where

$$\tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i}) \triangleq \frac{1}{2C_i} \sum_{\Omega_i} \frac{\omega(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)}{1 + \sum_{j=1, j\neq i}^{N} \frac{q_{jm}(\mathbf{x}_n)}{q_{im}(\mathbf{x}_n)} k(||\frac{\mathbf{x}_n - \mathbf{y}_j}{h_j}||^2)}. \quad (18)$$

This proximation means that the pixels in the occlusion regions are further down-weighted as

$$1/\left(1 + \sum_{j=1, j\neq i}^{N} \frac{q_{jm}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_j}{h_j}||^2)}{q_{im}(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)}\right) \rightarrow 1/\left(1 + \sum_{j=1, j\neq i}^{N} \frac{q_{jm}(\mathbf{x}_n)}{q_{im}(\mathbf{x}_n)} k(||\frac{\mathbf{x}_n - \mathbf{y}_j}{h_j}||^2)\right) \quad (19)$$

This is reasonable, since we don't explicitly recover the occlusion relations among the targets and a natural choice is to reduce their contributions to the weighted histograms.

4.2. Game construction and formulation

Although it is natural to design a game to model the competition among multiple trackers, the construction of the game cannot be arbitrary, *e.g.* based on intuitions or heuristics, because the equilibrium of the game may not necessarily be a solution to MTT. For example, if we formulate a naive non-cooperative game $\left[N, \{\mathbb{R}^2\}, \{\hat{\rho}_i(\mathbf{y}_i, \mathbf{y}_{-i})\}\right]$, where the players correspond to the individual trackers, the strategy for each player is the motion $\mathbf{y}_i \in \mathbb{R}^2$, and its utility $\hat{\rho}_i(\mathbf{y}_i, \mathbf{y}_{-i})$ is the generalized Bhattacharyya coefficient. This naive game is unable to assure a social optimal behavior (that corresponds to a good joint solution to MTT), because each tracker will try to solely increase its own utility. Special care has to be taken in the game construction.

A local optimum $\{\mathbf{y}_1^*, \cdots, \mathbf{y}_N^*\}$ of $J_3(\mathbf{y}_1, \cdots, \mathbf{y}_N) \triangleq r_{tot}(\mathbf{y}_1, \cdots, \mathbf{y}_N)$ is a good solution to MTT. The solution must satisfy the Karush-Kuhn-Tucker (KKT) conditions,

$$\frac{\partial r_{tot}(\mathbf{y}_1, \cdots, \mathbf{y}_N)}{\partial \mathbf{y}_i}|_{\{\mathbf{y}_1^*, \cdots, \mathbf{y}_N^*\}} = 0, \quad \forall i \leq N. \quad (20)$$

Thus, the N.E. of the game we construct must also satisfy these conditions. In view of this, we design a game $G = \left[N, \{\mathbb{R}^2\}, \{r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i})\}\right]$. At the N.E. $\{\mathbf{y}_1^*, \cdots, \mathbf{y}_N^*\}$ of this game, $\forall$ player i and its optimal strategy $\mathbf{y}_i^*$, we have $r_{tot}(\mathbf{y}_i^*, \mathbf{y}_{-i}^*) \geq r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i}^*), \forall \mathbf{y}_i$, by definition of N.E.. Since r_{tot} is continuous, $\nabla_{\mathbf{y}_i} r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i}^*)|_{\mathbf{y}_i^*} = 0, \forall i$, is held at N.E.. Consequently, the N.E. also satisfies the KKT conditions of J_3. Therefore, this construction of the game is plausible, and maximizing J_3 is equivalent to finding the N.E.. Fortunately, this can be solved efficiently by a decentralized best response updating, as described below.

4.3. Finding a Nash Equilibrium

To find a N.E., we design a decentralized synchronous scheme to update the best response for each tracker. Namely, $\forall i$, assuming all the other trackers' locations $\mathbf{y}_{-i}$ are given, we find the best $\hat{\mathbf{y}}_i$ that maximizes the utility $r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i})$, *i.e.* to solve $\nabla_{\mathbf{y}_i} r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i}) = 0$. The justification of this iterative process can be found in Sec. 5. We have, $\forall i$,

$$\nabla_{\mathbf{y}_i} r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i}) = \nabla_{\mathbf{y}_i} \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i}) + \sum_{j\neq i}^{N} \nabla_{\mathbf{y}_i} \tilde{r}_j(\mathbf{y}_j, \mathbf{y}_{-j}) = 0. \quad (21)$$

Eq. 21 can be solved in a closed-form. To make the derivation clear, we denote

$$\eta_{ii}(\mathbf{x}_n) \triangleq \frac{\omega_i(\mathbf{x}_n)}{1 + \sum_{j=1, j\neq i}^{N} \frac{q_{jm}(\mathbf{x}_n)}{q_{im}(\mathbf{x}_n)} k(||\frac{\mathbf{x}_n - \mathbf{y}_j}{h_j}||^2)}. \quad (22)$$

$$\eta_{ji}(\mathbf{x}_n) \triangleq \frac{\omega_j(\mathbf{x}_n) k(||\frac{\mathbf{x}_n - \mathbf{y}_j}{h_j}||^2)}{(1 + \sum_{l=1, l\neq j}^{N} \frac{q_{lm}(\mathbf{x}_n)}{q_{jm}(\mathbf{x}_n)} k(||\frac{\mathbf{x}_n - \mathbf{y}_l}{h_l}||^2))^2}, \quad (23)$$

Then, we have

$$\nabla_{\mathbf{y}_i} \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i}) = \frac{1}{C_i h_i^2} \sum_{\Omega_i} \eta_{ii}(\mathbf{x}_n) g(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)(\mathbf{x}_n - \mathbf{y}_i), \quad (24)$$

and for $i \neq j$, we have,

$$\nabla_{\mathbf{y}_i} \tilde{r}_j(\mathbf{y}_j, \mathbf{y}_{-j}) = -\frac{1}{C_j h_i^2} \sum_{\Omega_j \cap \Omega_i} \eta_{ji}(\mathbf{x}_n) g(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)(\mathbf{x}_n - \mathbf{y}_i) \quad (25)$$

Please note $\mathbf{y}_i$ merely influences $\tilde{r}_j(\mathbf{y}_j, \mathbf{y}_{-j})$ through the overlapped region $\{\mathbf{x}_n \in \Omega_j \cap \Omega_i\}$ and $g(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)$ is uniform for Epanechnikov kernel. $\nabla_{\mathbf{y}_i} \tilde{r}_j(\mathbf{y}_j, \mathbf{y}_{-j})$ acts as a force of the jth tracker that pushes away the ith tracker.

Plugging Eq. 24 and Eq. 25 to Eq. 21, we can solve the best $\hat{\mathbf{y}}_i$ given $\mathbf{y}_{-i}$ in a closed form. To make things clear,

we define two more coefficients $w_{ii}(\mathbf{x}_n)$ and $w_{ji}(\mathbf{x}_n)$ for pixel $\mathbf{x}_n \in \Omega_i$,

$$w_{ii}(\mathbf{x}_n) \triangleq \frac{1}{C_i h_i^2}\eta_{ii}(\mathbf{x}_n) g(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2), \forall \mathbf{x}_n \in \Omega_i, \tag{26}$$

$$w_{ji}(\mathbf{x}_n) \triangleq \begin{cases} -\frac{1}{C_j h_i^2}\eta_{ji}(\mathbf{x}_n) g(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2) & \mathbf{x}_n \in \Omega_i \cap \Omega_j \\ 0 & \mathbf{x}_n \notin \Omega_i \cap \Omega_j \end{cases}. \tag{27}$$

We have,

$$\nabla_{\mathbf{y}_i} r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i}) = \sum_{j=1}^{N} \nabla_{\mathbf{y}_i} \tilde{r}_j(\mathbf{y}_j, \mathbf{y}_{-j})$$

$$= \sum_{\Omega_i} \mathbf{x}_n \sum_{j=1}^{N} w_{ji}(\mathbf{x}_n) - \mathbf{y}_i \sum_{\Omega_i} \sum_{j=1}^{N} w_{ji}(\mathbf{x}_n) = 0. \tag{28}$$

Therefore, considering the interference of the target i to all the others targets and given the locations of other targets, the best $\hat{\mathbf{y}}_i$ that maximizes the utility is

$$\hat{\mathbf{y}}_i = \frac{\sum_{j=1}^{N} \sum_{\Omega_i} \mathbf{x}_n w_{ji}(\mathbf{x}_n)}{\sum_{j=1}^{N} \sum_{\Omega_i} w_{ji}(\mathbf{x}_n)}, \quad \forall i. \tag{29}$$

For each frame $I^{(t)}$, when N trackers approach to each other, we can iteratively update $\mathbf{y}_i, i = 1, \cdots, N$ by Eq. 29. This iterative process reaches an equilibrium that achieves a local optimum of the joint motion estimation.

A geometrical explanation is the following. We can view $\hat{\mathbf{y}}_i$ as a combination of forces $\hat{\mathbf{y}}_{i \leftarrow j}$ which is the solution to $\nabla_{\mathbf{y}_i} \tilde{r}_j(\mathbf{y}_j, \mathbf{y}_{-j}) = 0$ as

$$\hat{\mathbf{y}}_{i \leftarrow j} = \frac{\sum_{\Omega_i} \mathbf{x}_n w_{ji}(\mathbf{x}_n)}{\sum_{\Omega_i} w_{ji}(\mathbf{x}_n)}. \tag{30}$$

$\hat{\mathbf{y}}_{i \leftarrow j}$ acts as tracker j's counter force to tracker i when considering $\mathbf{y}_i$'s interference in $\tilde{r}_j(\mathbf{y}_j, \mathbf{y}_{-j})$. This can be visualized in Fig. 1.

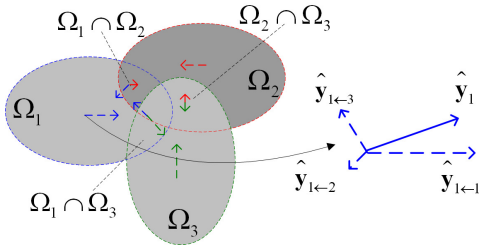

Figure 1. Illustration of force combination for $\hat{\mathbf{y}}_i$.

4.4. Algorithm summary

We summarize our game-theoretic MTT algorithm. If a subset of targets approach to each other, and their hypotheses are overlapped (the distances less than a threshold), we generate a game and use the algorithm in Fig. 3 to search for the N.E. If one target is isolated from others we use Mean-shift tracker. The procedure is summarized in Fig. 2.

Input : Frame $I^{(t)}$, target models $\{\mathbf{q}_i\}$, and initial states of the set of individual trackers $\theta^{(t-1)} = \{\mathbf{y}_i^{(t-1)}, h_i^{(t-1)}\}$ for $i = 1, \cdots, N'$.
Output: Tracking results $\theta^{(t)} = \{\mathbf{y}_i^{(t)}, h_i^{(t)}\}$ for $i = 1, \cdots, N'$.

1. Divide trackers into different groups if they are in close vicinity.
2. For each group of trackers, if it has more than one tracker in the group, generate a game and call the algorithm in Fig. 3, otherwise call Mean-shift tracker [2].
3. For each individual tracker, search $h_i^{(t)}$ with discrete scale factors $\{0.95, 1, 1.05\}$ to maximize its generalized Bhattacharyya coefficient $\hat{\rho}(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_{-i})$.

Figure 2. Procedure of game-theoretic MTT.

Input : Frame I, target models $\{\mathbf{q}_i\}$, and initial states of the set of individual trackers $\{\mathbf{y}_i^0, h_i\}$ for $i = 1, \cdots, N$.
Output: Target locations $\{\hat{\mathbf{y}}_i, i = 1, \cdots, N\}$ at the equilibrium.

1. For each tracker i, determine Ω_i and calculate $\hat{\mathbf{p}}_i(\mathbf{y}_i, \mathbf{y}_{-i})$ by Eq. 9.
2. In order to calculate $\nabla_{\mathbf{y}_i} \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})$ in Eq. 24, for each pixel $\mathbf{x}_n \in \Omega_i$, calculate
   - $\omega_i(\mathbf{x}_n)$ by Eq. 14,
   - $\eta_{ii}(\mathbf{x}_n)$ by Eq. 22,
   - $w_{ii}(\mathbf{x}_n)$ by Eq. 26.
3. In order to calculate $\nabla_{\mathbf{y}_j} \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})$ in Eq. 25 (note switch subscript i and j), for tracker $j \neq i, \Omega_i \cap \Omega_j \neq \emptyset$, for each pixel $\mathbf{x}_n \in \Omega_i \cap \Omega_j$, calculate
   - $\eta_{ij}(\mathbf{x}_n)$ according to Eq. 23,
   - $w_{ij}(\mathbf{x}_n)$ according to Eq. 27.
4. For tracker i, calculate $\hat{\mathbf{y}}_i$ given $\mathbf{y}_{-i}$ by Eq. 29.
5. If all $\{\hat{\mathbf{y}}_i \ \forall i = 1, \cdots, N\}$ are stationary, exit; otherwise go to Step 1.

Figure 3. Algorithm for finding N.E. in game-theoretic MTT.

5. Game theoretic analysis

In the game G we have constructed, the utility function of each player is the joint matching $r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i}) = \sum_i^N \tilde{r}(\mathbf{y}_i, \mathbf{y}_{-i})$, which forces an individual tracker to take other trackers' influences into consideration rather than only focusing on its own interest. $\nabla_{\mathbf{y}_i} \tilde{r}_j(\mathbf{y}_j, \mathbf{y}_{-j})$, *i.e.* the sensitivity of tracker j's matching w.r.t tracker i's motion $\mathbf{y}_i$, can be regarded as a price tracker j charges tracker i and counter reacts to $\mathbf{y}_i$ through $\hat{\mathbf{y}}_{i \leftarrow j}$.

To analyze whether the Nash Equilibrium can be achieved by the best response updating for game $G = \left[N, \{\mathbb{R}^2\}, \{r_{tot}(\mathbf{y}_1, \cdots, \mathbf{y}_N)\}\right]$, we resort to the following definition and theorem in the supermodular game theory [15, 16].

Definition 1 *A game $G = \{N, S, \{f_i\}\}$ is a supermodular (submodular) game if the set S of feasible joint strategies is a sublattice, and each utility function f_i is supermodular (submodular) function on S.*

Theorem 1 *In a supermodular (submodular) game $G = \{N, S, \{f_i\}\}$, (a) there exists at least one Nash Equilibrium; (b) if each player starts from any feasible strategy and uses best response updating, then the joint strategies will eventually converge to a Nash Equilibrium.*

For details about supermodular games, we refer the readers to Chapter 4 in [15] and Chapter 7 in [16].

Based on the supermodular game theory, to show the best response updating can reach a N.E., a sufficient condition includes 1) the solution of Eq. 21 is a best response of $\hat{\mathbf{y}}_i$ given fixed $\mathbf{y}_{-i}$, and 2) the game G is a supermodular/submodular game. Condition 1 is satisfied since $r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i})$ is concave on $\mathbf{y}_i$ in that the Epanechnikov kernel function k is non-negative and strictly concave. The details are given in Appendix A. The condition 2 can be satisfied in certain $\Omega_i, i = 1, \cdots, N$ where each utility function is submodular function, which is given in Appendix B.

6. Experiments

We demonstrate the proposed game-theoretic MTT by using both synthesized and real video (downloaded from *Google Video*). The basic individual tracker is a Mean-shift trackers with 32×32 2D histogram in the Hue-Saturation space. To purely evaluate the performance of the proposed method, we do not incorporate motion dynamic prior, object detectors, and background subtraction, although it is easy to incorporate them. The method is implemented in C++ and tested on Pentium IV 3Ghz PC. Empirically, the best response updating converges very quickly within 3-10 iterations, so the computations are almost the same as that in multiple independent Mean-shift trackers.

6.1. Example of best response updating

First, we show an example of the best response updating for tracking the hands and the face in a sign language video. The first 4 images in Fig. 4 show the positions of the hands and the face at the first 3 iterations and at the last iteration during the best response updating. We observe that the sum of generalized Bhattacharyya coefficients $\sum_{i=1}^3 \hat{\rho}(\mathbf{y}_i, \mathbf{y}_{-i})$ monotonically increases as shown in the last graph. But the individual $\hat{\rho}(\mathbf{y}_i, \mathbf{y}_{-i})$ may be up and down. This is a rather difficult case because the hands and the face share the same skin tones. In our method, the competition ends up at an equilibrium that gives a good estimation of them.

6.2. Synthesized video

We synthesize two videos in which there are 3 different targets and 5 identical targets, respectively. The backgrounds include random noise and 10-20 small targets that are wandering randomly. Frame samples are shown in Fig. 5. The trackers are drawn in different colors and a red dash ellipse indicates the group of trackers that are engaged in the game. The final motion $\hat{\mathbf{y}}_i$ are drawn at the centers of the targets. From the test results, the competition among the targets leads to an equilibrium and largely avoids the coalescence problem.

6.3. Real video

We further test the proposed approach in real sign language and sports videos. These are very challenging tests for MTT. The hand gesturing in sign language video (Fig. 6) is fast and the hand shape is deformable. Since the color of the hands and the face are quite similar, when the hands moving in front of the head, it is very likely that independent trackers will fail as shown in the 2nd row of Fig. 6. On the contrary, in our method, the interference from the face tracker to the hands tends to push the hands away from the face, which greatly alleviates coalescence phenomenons.

Sports video is another large category where the athletes generally wear similar sports suits and may have very complicated interactions. Therefore tracking the people in sports video is a very difficult task. We show the tracking results for `kid soccer`, `free style soccer` and `volleyball`. The proposed method can follow the people with complicated occlusions. The comparison to the results of independent trackers are in the supplemental materials.

7. Conclusion

In this paper we introduce a new view of game theory to the study of multiple target tracking. The competition of individual trackers is formulated as a game and we bridge the solution to the joint motion estimation and the Nash Equilibrium of the game. Consequently, the maximization of the

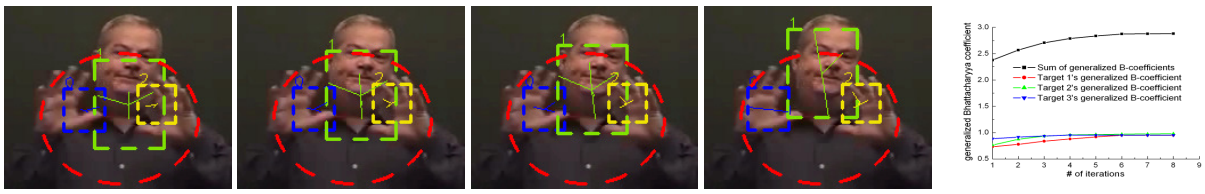

Figure 4. Illustration of best response updating procedure: iteration #0, 1, 2, and 8.

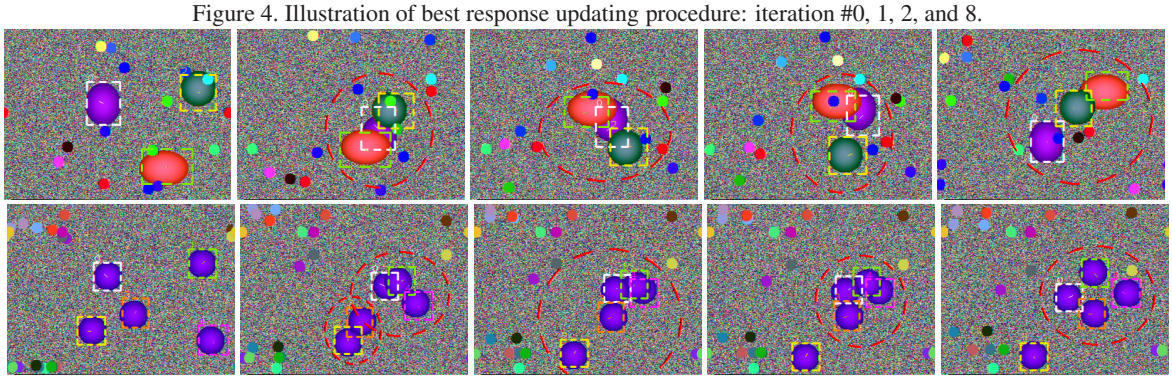

Figure 5. Tracking synthesized video: (1st row) 3 different targets for frame #1, 15, 42, 427, and 500; (2nd row) 5 identical targets for frame #1, 13, 19, 20, 25.

joint likelihood can be decentralized. The N.E. of this game can be solved by an efficient iterative procedure in a closed form. The proposed method achieves promising results in tracking quasi-identical targets in both synthesized and real video sequences. The future work includes the incorporation of motion dynamic models in the trackers' utilities and faster algorithms for computing approximate N.E.

Acknowledgments

This work was supported in part by National Science Foundation Grants IIS-0347877 and IIS-0308222.

Appendix

A. Proof of Eq. 29 is a best response

To show Eq. 29 is the best response of $\hat{\mathbf{y}}_i$ given fixed $\mathbf{y}_{-i}$, we need to show the solution $\hat{\mathbf{y}}_i$ of Eq. 21 is a global optimum of $r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i})$. We prove this by showing $r_{tot}(\mathbf{y}_i, \mathbf{y}_{-i}) = \sum_{j=1}^{N} \tilde{r}_j(\mathbf{y}_1, \cdots, \mathbf{y}_N)$ is concave.

Denote $\mathbf{y}_i = \{u_i, v_i\}$, given $\mathbf{y}_{-i}$ are fixed, $\tilde{r}_i(\mathbf{y}_i) = \tilde{r}_i(u_i, v_i)$ and $\tilde{r}_j(\mathbf{y}_i) = \tilde{r}_j(u_i, v_i)$. Note $g(||\frac{\mathbf{x}_n - \mathbf{y}_i}{h_i}||^2)$ is positive and uniform for Epanechnikov kernel. From Eq. 24 and Eq. 25, we have

$$\frac{\partial \tilde{r}_i(u_i, v_i)}{\partial u_i \partial v_i} = 0, \frac{\partial \tilde{r}_i(u_i, v_i)}{\partial u_i \partial u_i} = \frac{\partial \tilde{r}_i(u_i, v_i)}{\partial v_i \partial v_i} = -\sum_{\Omega_i} w_{ii}(\mathbf{x}_n).$$

$$\frac{\partial \tilde{r}_j(u_i, v_i)}{\partial u_i \partial v_i} = 0, \frac{\partial \tilde{r}_j(u_i, v_i)}{\partial u_i \partial u_i} = \frac{\partial \tilde{r}_j(u_i, v_i)}{\partial v_i \partial v_i} = -\sum_{\Omega_i} w_{ji}(\mathbf{x}_n).$$

So in the Hessian matrix of $\sum_{j=1}^{N} \tilde{r}_j(u_i, v_i)$, the elements on the diagonal are $-\sum_{j=1}^{N} \sum_{\Omega_i} w_{ji}(\mathbf{x}_n)$ and 0 for elements off the diagonal, it is negative definite which indicates it is concave over $\mathbf{y}_i = \{u_i, v_i\}$.

B. Conditions for G being a submodular game

To show a game is supermodular (submodular) game we need to show the joint strategy space is defined on a sublattice and all utility functions are supermodular (submodular) functions on the joint strategy space. Any non-empty compact subset of $\mathbb{R}^n$ is a sublattice of $\mathbb{R}^n$ [16]. So the first requirement is satisfied in our game G. For the second condition, we have this theorem [16]:

Theorem 2 *Let $X \subset \mathbb{R}^n$ and $f : X \to \mathbb{R}$. The function f is supermodular iff it satisfies increasing (decreasing) differences on X. If f is twice differentiable, f is supermodular iff $\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0$, or submodular iff $\frac{\partial^2 f}{\partial x_i \partial x_j} \leq 0$, $\forall i, j$.*

Thus, denote $\mathbf{y}_i = \{u_i, v_i\}$, we need to examine $\frac{\partial \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})}{\partial u_i \partial u_j}$, $\frac{\partial \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})}{\partial v_i \partial v_j}$, $\frac{\partial \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})}{\partial u_i \partial v_j}$, and $\frac{\partial \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})}{\partial v_i \partial u_j}$ for $i \neq j$. In addition, we need to check $\frac{\partial \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})}{\partial u_k \partial u_l}$, $\frac{\partial \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})}{\partial v_j \partial v_l}$, $\frac{\partial \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})}{\partial u_j \partial v_l}$, and $\frac{\partial \tilde{r}_i(\mathbf{y}_i, \mathbf{y}_{-i})}{\partial v_j \partial u_l}$ for $j, l \neq i$. Whether these conditions hold depends on the $\{\Omega_i, i = 1, \cdots, N\}$ and can be checked analytically. We observe the constructed game G is submodular when the occlusion regions are small and the kernel centers are not occluded. Due to the page limit, we are unable to list the derivation of each term, these conditions can be checked as a by-product in best response updating given $\{\Omega_i, i = 1, \cdots, N\}$.

References

[1] Yaakov Bar-Shalom and Thomas E. Fortmann. *Tracking and Data Association*. Academic Press, 1988. 2

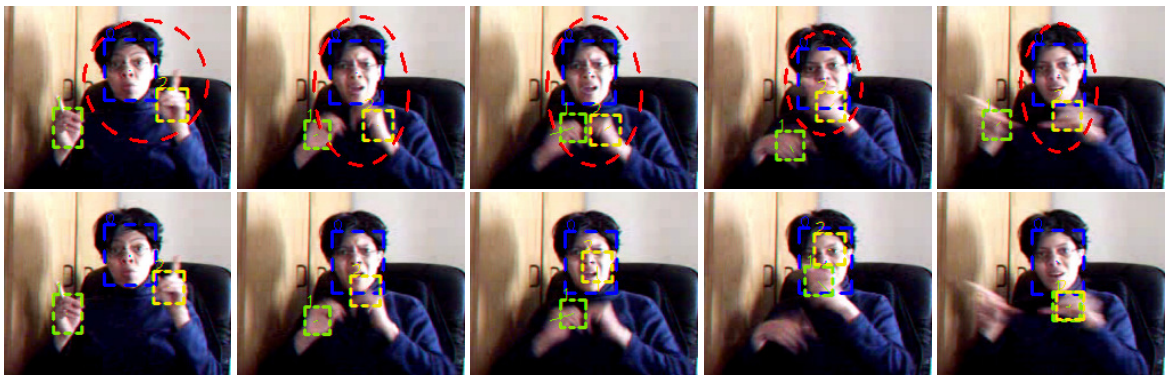

Figure 6. Tracking [`sign language`] for frame #1, 171, 172, 305, and 325, (1st row) game-theoretic MTT trackers and (2nd row) multiple independent trackers.

Figure 7. (1st row) tracking [`kid soccer`] for frame #40, 64, 79, 101, 109; (2nd row) tracking [`free style soccer`] for frame #1, 100, 250, 280, and 300; (3rd row) tracking [`volleyball`] for frame #1, 15, 40, 50, and 120.

[2] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Real-time tracking of non-rigid objects using mean shift. In *CVPR'00*, volume 2, pages 142–149, Hilton Head Island, South Carolina, June 13-15, 2000. 1, 2, 5

[3] Gregory D.Hager, Maneesh Dewan, and Charles V. Stewart. Multiple kernel tracking with SSD. In *CVPR'04*, volume 1, pages 790 – 797, Washington, DC, Jun.27-Jul.2 2004. 1

[4] Mei Han, Wei Xu, Hai Tao, and Yihong Gong. An algorithm for multiple object trajectory trackingt. In *CVPR'04*, volume 1, pages 864 – 871, Washington, DC, Jun.27-Jul.2 2004. 2

[5] Yan Huang and Irfan Essa. Tracking multiple objects through occlusions. In *CVPR'05*, volume 2, pages 1051 – 1058, San Diego, CA, June 20-25, 2005. 2

[6] Carine Hue and Jean-Pierre Le Cadre. Sequential monte carlo methods for multiple target tracking and data fusion. *IEEE Trans. Signal Processing*, 50(2):309 – 325, February 2002. 1, 2

[7] Michael Isard and John MacCormick. Bramble: A bayesian multiple-blob tracker. In *ICCV'01*, volume 2, pages 34–41, Vancouver, Canada, July 7-14, 2001. 1, 2

[8] Zia Khan, Tucker Balch, and Frank Dellaert. MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Trans. Pattern Anal. Machine Intell.*, 27(11):1805 – 1819, November 2005. 1, 2

[9] John MacCormick and Andrew Blake. A probabilistic exclusion principle for tracking multiple objects. In *ICCV'99*, pages 572 – 578, Corfu, Greece, 21-22 1999. 1, 2

[10] John Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295, 1951. 1, 2

[11] Kenji Okuma, Ali Taleghani, Nando De Freitas, James J. Little, and David G. Lowe. A boosted particle filter: Multitarget detection and tracking. In *ECCV'04*, pages 28 – 39, Prague, Czech Republic, May11-14 2004. 1, 2

[12] Chrisopher Rasmussen and Gregory D. Hager. Probabilistic data association methods for tracking complex visual objects. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(6):560 – 576, June 2001. 1, 2

[13] Donald B. Reid. An algorithm for tracking multiple targets. *IEEE Trans. Automat. Contr.*, 24(6):843 – 854, December 1979. 1, 2

[14] Hai Tao, Harpreet S. Sawhney, and Rakesh Kumar. A sampling algorithm for detecting and tracking multiple targets. In *ICCV'99 Workshop on Vision Algorithms: Theory and Practice*, pages 53 – 58, Corfu, Greece, 21-22 1999. 1, 2

[15] Donald M. Topkis. *Supermodularity and Complementarity*. Priceton University Press, 1998. 2, 6

[16] Rakesh V. Vohra. *Advanced Mathematical Economics*. Routledge, 2005. 6, 7

[17] Bo Wu and Ram Nevatia. Tracking of multiple, partially occluded humans based on static body part detection. In *CVPR'06*, volume 1, pages 951 – 958, NYC, June 17-22, 2006. 2

[18] Ting Yu and Ying Wu. Collaborative tracking of multiple targets. In *CVPR'04*, volume 1, pages 834 – 841, Washington, DC, Jun.27-Jul.2 2004. 2

[19] Tao Zhao and Ram Nevatia. Tracking multiple humans in crowded environment. In *CVPR'04*, volume 2, pages 1063 – 1069, Washington, DC, Jun.27-Jul.2 2004. 1, 2