

Chapter 5

Coalitional Tracker for Deception Detection in Thermal Imagery

Jonathan Dowdall, Ioannis Pavlidis, and Panagiotis Tsiamyrtzis

Abstract We propose a novel tracking method that uses a network of independent particle filter trackers whose interactions are modeled using coalitional game theory. Our tracking method is general; it maintains pixel-level accuracy, and can negotiate surface deformations and occlusions. We tested our method in a substantial video set featuring nontrivial motion from over 40 objects in both the infrared and visual spectra. The coalitional tracker demonstrated fault-tolerant behavior that far exceeds the performance of single-particle filter trackers. Our method represents a shift from the typical tracking paradigms and may find application in demanding imaging problems across the electromagnetic spectrum.

Keywords: Tracking · Particle filter · Coalitional game theory · Thermal imaging

5.1 Introduction

The extraction of high-level information from video through the use of computer vision algorithms has become increasingly important over the past decade. A diverse array of applications use this technology, including quality control in the manufacturing sector [1, 2], surveillance in the security industry [3, 4], biomedical measurements for health care [5–7], and behavioral analysis [8–10]. Of key importance to all these computer vision applications is the ability to detect and track objects in their respective input video streams. The problem of tracking can be cast as guessing how things change over time. Specifically, tracking involves modeling how the parameters of the object modulate in successive input frames by using prior knowledge. When this is done accurately, it can be useful in a number of applications for which knowing the current state of a given target object is important. An intriguing line of computer vision research focuses on measurements of physiological signals on facial tissue. The measurements are performed on infrared imagery and are used

in biomedical [5–7] and behavioral applications [8–10]. Although a large body of work has been devoted to facial tracking research [11–13], we found the existing methods insufficient to achieve the high degree of accuracy required in imaging measurements of facial tissue. This was our initial motivation for exploring a novel tracking paradigm.

5.1.1 *Prior Work*

Computer vision tracking has been dominated by sequential Monte Carlo methods (particle filtering) [14] for the last several years. Among the most popular particle filter tracking methods is the CONDENSATION algorithm, which was introduced by Isard et al. circa 1998 [15–17].

An interesting tracking methodology based on deformable templates was also developed in parallel. Typical deformable templates focus on tracking object contours, not surfaces [18]. Therefore, they cannot adequately address out-of-plane tracking, like the case of left-right facial rotation.

Alternative tracking methodologies employ specific models of the target to provide better accuracy [19–21]. Unfortunately, this increased accuracy comes at the expense of speed and generality. A noteworthy modeling approach is known as active appearance modeling, and it takes into account both shape and texture [22, 23]. For example, Dornaika et al. [23] first recovered the three-dimension (3D) head pose using a deformable wire frame and then local motion associated with some facial features using active appearance model search. Such 3D active appearance models can potentially perform quality tracking in demanding facial-imaging applications in the visual spectrum. However, their performance may break in thermal infrared imagery due to thermal diffusion and the resulting fuzzy image edges. In such an environment, appearance models may have hard time maintaining 3D-2D (two-dimensional) correspondences, which are partly based on thermal gradients.

Tracking in the thermal infrared spectrum is of particular interest to us because recent research demonstrated that many vital signs, including blood flow [5], pulse [6], and breathing function [7], can be measured in this modality. The success of these measurements depends strongly on a reliable tracking method to register the motion of facial tissue.

Our method aims to achieve what sophisticated modeling methods reportedly achieve, but it is more general and robust. It does not employ a single explicit 3D model but many generic and cooperating 2D particle filter trackers, which are spatially distributed over the target's surface. Our effort can be seen as a first step toward developing a tracking methodology that is able to accurately track a wide array of targets across imaging modalities.

There has been some other work on multiple trackers that work together to follow *multiple objects* [24–27]. In contrast, we employ multiple trackers to track a *single object*.

5.2 Tracking Methodology

Our goal is to develop a general tracking methodology that can accurately monitor the motion of the target's surface even in the presence of deformation or partial occlusion. Many existing general tracking methods monitor the target's outline (not surface). This is a different and far easier problem.

We arrived at a fault-tolerant surface-tracking method that works on both infrared and visual video without resorting to explicit modeling. It uses a network of particle filter trackers that influence each other (see Fig. 5.1). Each individual tracker is unreliable at times, but the combination of many neighboring trackers produces robust performance. The intertracker influence is modeled as a coalitional game in which each tracker is a player, and the goal of the game is to propagate one's influence in subsequent frames of video. Within this framework, the winning coalition of trackers is used to calculate the state of the tracked object.

5.2.1 Tracking Network

We use a network of trackers to achieve accurate surface tracking and fault tolerance. Tracking is maintained even if all but one of the trackers fail in the tracking network. The trackers are each assigned a different portion of the target's surface to track (see Fig. 5.2). By default, the trackers are configured in a regular grid, although alternative configurations are possible through a feature selection mechanism. Intertracker communication allows trackers that are correctly tracking the target to



Fig. 5.1 Overview of the tracking method. Initialization consists of the following steps: **a** extraction of the user-selected region of interest from the input video, **b** subdivision of the Region of interest into the tracking network, and **c** individual tracker template creation. Tracking: the individual trackers in the tracking network follow their targets. Target state estimation consists of the following steps: **d** the winning coalition is produced, **e** the deformation mesh is calculated from the winning coalition, and **f** the deformation mesh is used to calculate the target state. The method proceeds from initialization to tracking (*arrow 1*) to target state estimation (*arrow 2*) and back to tracking (*arrow 3*). In the latter transition, the winning coalition is passed back to the tracking stage to distribute the intertracker influence.

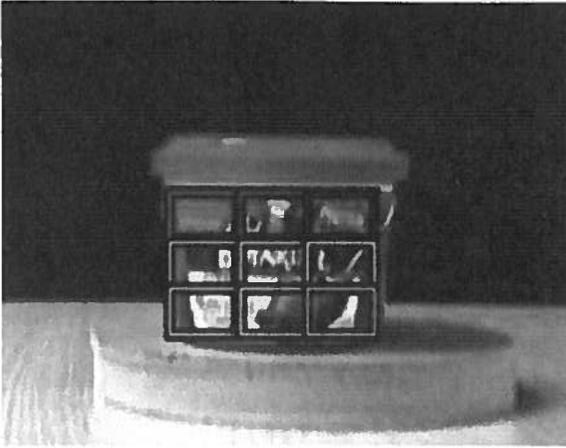


Fig. 5.2 Example of a 3×3 tracking network on a visual image. Each tracker in the network is shown in a different color. Each tracker is tracking a separate part of the target

“tip-off” other trackers that have become lost regarding the true location of the target’s surface. This intertracker influence is realized within a statistical framework and is managed by the coalitional game model described in Section 5.2.2.

The idea arose naturally in the effort to address the problem of facial tissue tracking in the infrared. As the subject’s head moves (e.g., left and right), part of the facial surface is occluded at times. Trackers that correspond to the occluded part of the face are aided by trackers that correspond to the exposed part.

In our implementation, each tracker in the tracking network is an individual particle filter tracker. We denote the state of each individual tracker i at time t by $x_i^{(t)}$ and its associated image observation by $z_i^{(t)}$. The target tracker’s prior will be formed using intrasamples $s_{(i,i)}^{(t)}$ from tracker i and intersamples $s_{(i,j)}^{(t)}$ that correspond to the (intertracker) influence of tracker i from tracker j . The intersamples $s_{(i,j)}^{(t)}$ are generated based on the initial relationship between the trackers involved in the exchange:

$$s_{(i,j)}^{(t)} = T_{x_i^{(0)}}^{x_j^{(0)}} x_j^{(t)} \quad (5.1)$$

where $s_{(i,j)}^{(t)}$ is the intersample generated by tracker j for tracker i , and $T_{x_i^{(0)}}^{x_j^{(0)}}$ is the transformation that gives a sample for tracker i given a state for target j at time t . The transformation $T_{x_i^{(0)}}^{x_j^{(0)}}$ is computed during initialization for every possible tracker pair $(x_j^{(0)}, x_i^{(0)})$.

5.2.2 The Coalitional Game

The tracking network is a versatile architecture for tracking objects, but it does not have any intrinsic method to generate the final target state or to manage tracker interaction. The simplest solution would be to allow every tracker to influence all of the other trackers. Unfortunately, this is not an optimal solution because trackers that have lost their target would be allowed to influence other trackers in the network that have not gone awry. This also highlights the problem of determining which trackers in the network are correctly tracking their targets and which ones have strayed away. What is needed is a mechanism that can determine the validity of each of the trackers, compute the target's state vector based on the valid trackers, and finally propagate the influence of the valid trackers to keep the network correctly tracking the target surface.

There are many optimization algorithms one can use to manage the network of trackers. We chose to optimize tracker interaction using a game theoretic solution for two main reasons: It naturally fits the problem space, and it is relatively simple. Game theory [28–31] has been successfully used to analyze topics ranging from simple deterministic games, to complex economic models [32, 33], and even to international negotiations [34, 35]. Our adaptation was to view the trackers as players in a cooperative game [36, 37] in which the objective was to increase their influence by forming coalitions with other trackers. The winning coalition would then be used to compute the state vector of the target and subsequently propagate its influence onto the entire tracking network.

Specifically, the members $m_j^{(i)}$ of the winning coalition C' influence every other tracker i in the tracking network by adding intersamples $s_{(i,j)}^{(i)}$. Trackers that are not members of the winning coalition cannot propagate any influence at all. The intuitive affinity of the problem space to cooperative gaming is apparent in the example of facial tissue tracking. There, the winning coalition is composed mostly of trackers that correspond to the exposed part of the face. These are trackers that feature high-quality information and give a “helping hand” (influence) to the “clueless” trackers that correspond to the occluded or deformed part of the face.

The coalitional form of an N -tracker game is given by the pair (Ω, Π) , where $\Omega = \{1, 2, \dots, N\}$ is the set of trackers and Π is a real-valued function, called the *characteristic function* of the game, defined on the set of all coalitions (subsets of Ω), which has cardinality 2^N and satisfying $\Pi(\emptyset) = 0$ [28]. In other words, the empty set has value zero. The size of a coalition C will be denoted from now on by k , where $k \in \{1, 2, \dots, N\}$, and there are $\binom{N}{k}$ coalitions of size k . The quantity $\Pi(C_k)$ may be considered as the value, or worth, or power, of coalition $C_k \subset \Omega$ when its members act together as a unit.

The definition of a coalitional game is quite general and leaves the specification of the characteristic function to the game designer. We designed a characteristic function for the tracking game that encompasses four scores. These scores are calculated from the trackers participating in the coalition under consideration at time t :

- template match $\alpha^{(t)}$
- geometric alignment $\beta^{(t)}$
- interframe projection agreement $\gamma^{(t)}$
- interframe membership retention $\delta^{(t)}$

The characteristic scores support the fact that quality tracking is characterized by consistency in the content and geometric configuration of the individual trackers. Specifically, the template match score rewards trackers that maintain consistent imaging content. The geometric alignment score favors coalitions whose members have geometric alignment analogous to the original ($t=0$) configuration. The interframe projection agreement score is a continuity constraint. It improves robustness by penalizing abrupt (and improbable) changes of the projected state of the target between successive frames. The interframe membership retention score is also a continuity constraint. It reflects the tendency of the winning coalition from the previous time step to retain its members.

The template match score $\alpha_{C_k}^{(t)}$ for a coalition C_k of size k at time t is given by

$$\alpha_{C_k}^{(t)} = \frac{1}{k} \sum_{i=1}^k \alpha_{m_i}^{(t)} \quad (5.2)$$

where $\alpha_{m_i}^{(t)}$ refers to the template match score (a number in $[0,1]$) of member m_i in the coalition C_k at time t .

For the second and third scores, we first need to define the function that measures the geometric alignment between two target projections (see Fig. 5.3), as are computed from samples s_i and s_j :

$$G(\mathcal{F}(s_i), \mathcal{F}(s_j)) = G(S_i, S_j) = \omega \times \left[1 - \frac{\sqrt{(S_{ix} - S_{jx})^2 + (S_{iy} - S_{jy})^2}}{M_d} \right] \\ + (1 - \omega) \times \left[1 - \frac{|S_{i\theta} - S_{j\theta}|}{M_\theta} \right] \quad (5.3)$$

where $\mathcal{F}(s)$ is a function that transforms the tracker sample s into its corresponding target projection S , (S_{ix}, S_{iy}) are the (x, y) coordinates of the center of target projection S_i , while $S_{i\theta}$ is the angle of rotation about the center of target projection S_i . M_d is set to the maximum movement allowed by the target in a single frame, while M_θ is the (positive) maximum rotation allowed by the target in a single frame. The weight ω appropriately penalizes the center and angle discrepancies. Ideally, the target projections in Fig. 5.3 c should have coincided (perfect alignment), so that the combined projection of the two tracker samples is reminiscent of the original target shape. Note that the upper bound for $G(\cdot, \cdot)$ is 1 (when the two target projections are identical), but the lower bound is not necessarily 0. This would have been the case if we chose M_d and M_θ to be the maximum observed values at time t , but this would

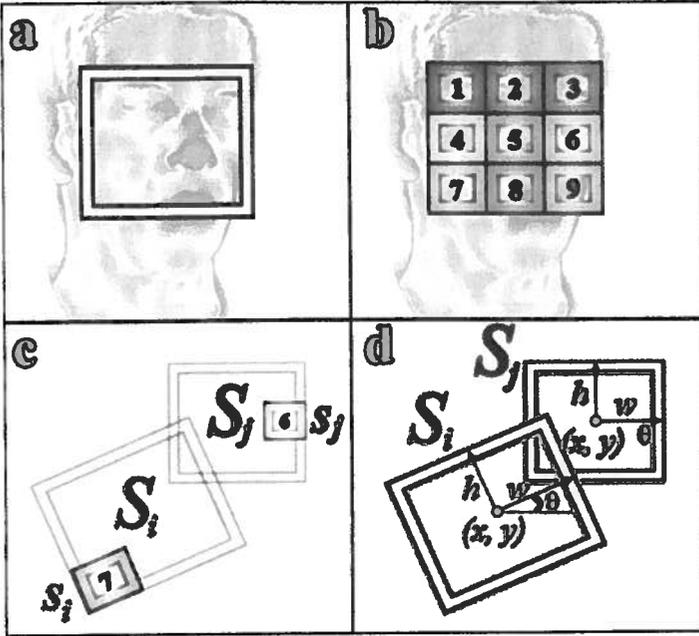


Fig. 5.3 Geometric alignment of tracker's target projections: a target projection at $t = 0$, b tracker network overlaid on the initial target projection, c trackers 6 and 7 at a subsequent time along with their corresponding target projections, d parameterization of target projections to facilitate measurement of geometric alignment

have slowed the computation. Besides, we do not mind giving negative scores to some tracker pairs (i.e., penalizing as opposed to rewarding them) whose geometric alignment is very bad.

Having defined the geometric alignment function for a pair of samples [see Eq. (5.3)], we use it to compute the geometric alignment score $\beta_{C_k}^{(t)}$ of a coalition of size k :

$$\beta_{C_k}^{(t)} = \frac{f(k)}{\binom{k}{2}} \sum_{i=1}^{k-1} \sum_{j=i+1}^k G(S_i^{(t)}, S_j^{(t)}) \quad (5.4)$$

where $S_i^{(t)}$ and $S_j^{(t)}$ are target projections corresponding to the samples with the highest template match scores for coalition members m_i, m_j , respectively. Regarding the function $f(k)$, we have $f(1) = 0$, and it is nondecreasing for $k = 2, 3, \dots, N$. The $\beta_{C_k}^{(t)}$ is analogous to the average of the geometric alignment of all possible tracker pairs in the coalition. In general, as the size of the coalition increases, the average of the geometric alignment function of the members of the coalition decreases. To compensate for that loss, we introduced the linear function $f(k)$, whose role is to reward higher-order coalitions as opposed to lower-order ones.

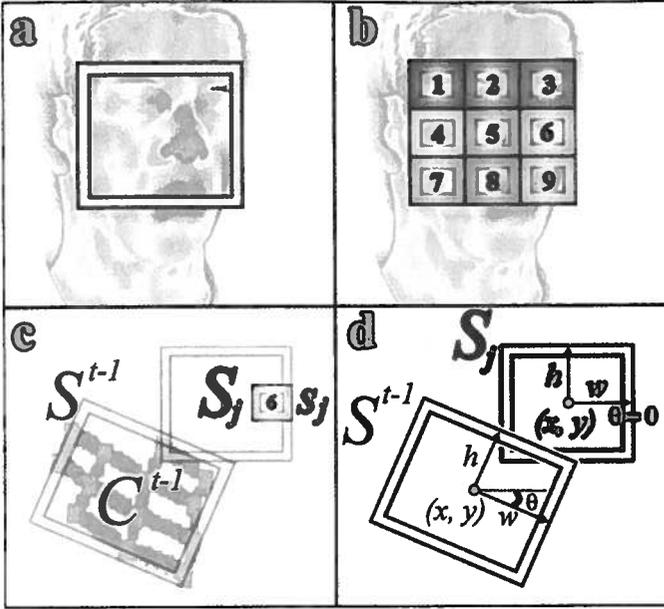


Fig. 5.4 Interframe projection agreement: **a** target projection at $t = 0$, **b** tracker network overlaid on the initial target projection, **c** the target projection at time $t - 1$ and tracker 6 with its corresponding target projection at time t , **d** parameterization of target projections to facilitate measurement of interframe projection agreement

We also use the geometric alignment function for a pair of samples [see Eq. (5.3)] to compute the interframe projection agreement $\gamma_{C_k}^{(t)}$ score (see Fig. 5.4):

$$\gamma_{C_k}^{(t)} = \frac{1}{k} \sum_{i=1}^k G \left(S_i^{(t)}, S^{(t-1)} \right) \quad (5.5)$$

where $S_i^{(t)}$ is the target projection corresponding to the sample with the highest template match score for coalition member $m_i^{(t)}$ at time t ; $S^{(t-1)}$ is the target projection corresponding to the target state at time $t - 1$ (previous frame). The interframe membership retention score $\delta_{C_k}^{(t)}$ for a coalition C_k of size k at time t is given by

$$\delta_{C_k}^{(t)} = \frac{1}{k} \sum_{i=1}^k \Delta \left(m_i^{(t)}, C^{t-1} \right) \quad (5.6)$$

where m_i^t is the i th member of coalition C_k at time t , C^{t-1} is the winning coalition from the previous time step, and Δ is defined as

$$\Delta(m, C) = \begin{cases} -1 & \text{if } m \text{ is not a member of } C \\ +1 & \text{if } m \text{ is a member of } C \end{cases} \quad (5.7)$$

where m is a tracker, and C is a coalition.

Having defined the four scores, we proceed with the definition of the characteristic game function $\Pi^{(t)}(C_k)$:

$$\Pi^{(t)}(C_k) = \omega_\alpha \times \alpha_{C_k}^{(t)} + \omega_\beta \times \beta_{C_k}^{(t)} + \omega_\gamma \times \gamma_{C_k}^{(t)} + \omega_\delta \times \delta_{C_k}^{(t)} \quad (5.8)$$

where $\omega_\alpha, \omega_\beta, \omega_\gamma, \omega_\delta$ are the weights (values range in $[0,1]$ and sum to 1) assigned to the four scores. Note that because of the function $f(k)$ in the geometric alignment score, the characteristic score may exceed the value of 1. This may happen when we have higher-order coalitions and quite good geometric alignment.

For every size of coalition $k \in \{1, 2, \dots, N\}$, we have $\binom{N}{k}$ different coalitions of size k , out of which we select the one with the highest payoff. To avoid complicating the symbology, C_k will continue to denote the preferred coalition of size k . Thus, we decide the winning coalition C^t at time t to be

$$C^t = \arg \max_{C_k} \Pi^{(t)}(C_k) \quad (5.9)$$

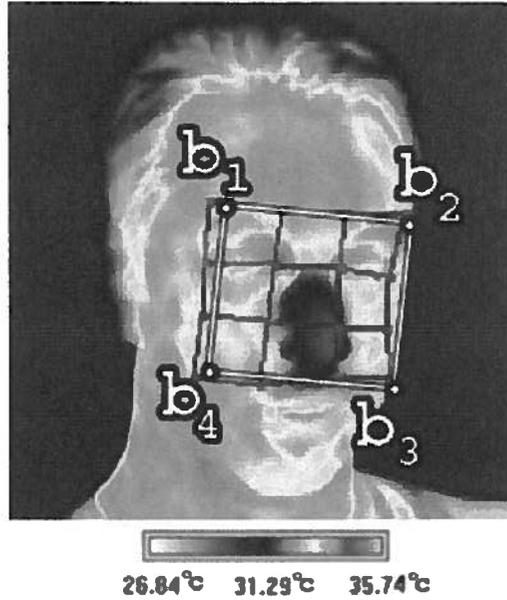
In coalitional game theory, sometimes the characteristic function is a nondecreasing function of the size of the coalition (i.e., superadditivity) [28]. In our case, this is not desirable because there may exist trackers that have lost their targets. In other words, the grand coalition (i.e., the coalition where all players/trackers participate) is not always the optimal to use. Thus, we need to give rewards to coalitions in such a way that the winning coalition is the coalition whose members best approximate the target. This is achieved by reducing the characteristic function of the coalition if it acquires poor trackers. Superadditivity is also related to the $f(k)$ function since if $f(k)$ increases, say exponentially, then the geometric alignment score will dominate the other three scores, allowing superadditivity. In our case, having a linear $f(k)$ worked fairly well.

5.2.3 Target State Estimation

We compute the final target state S^t from the winning coalition C^t in two steps. In the first step, we compute the deformation mesh M^t from the winning coalition. The deformation mesh M^t is composed of a set of points $A = (a_1, \dots, a_m)$, which are distributed over the selected target region during the initialization step. Each point is linked to anywhere between one through four trackers depending on its spatial location; one on the corners, two on the borders, and four on the inside. For each point, a transformation matrix $T_{a_i}^{c_j}$ is computed that, when applied to the center c_j of tracker j , gives the location of the point a_i :

$$a_i = \frac{1}{\sum_{j=1}^{n_i} \omega_j} \sum_{j=1}^{n_i} \begin{bmatrix} c_{jx} \\ c_{jy} \end{bmatrix} T_{a_i}^{c_j} \omega_j \quad (5.10)$$

Fig. 5.5 Border points of the target projection. The target projection is shown in white. The deformation mesh is shown in blue, and the deformation mesh points are shown



where a_i is one of the points in the deformation mesh M^t , n_i is the number of trackers linked to the mesh point a_i , and ω_j is the weight associated with tracker j . If the tracker j is a member of the winning coalition, the associated weight is the tracker's template match score ($\omega_j = \alpha_m^{(t)}$); otherwise, it is 0 $\omega_j = 0$. Next, the four border points outlining the target projection in the clockwise direction $B = (b_1, \dots, b_4)$ are computed from the mesh points A (see Fig. 5.5).

$$b_i = \frac{1}{\omega_{tot}} \sum_{j=1}^m \begin{bmatrix} a_{jx} \\ a_{jy} \end{bmatrix} T_{b_i}^{a_j} \omega_{a_j} \quad (5.11)$$

where m is the number of points in the deformation mesh, b_i is one of the border points, a_j is one of the points in the deformation mesh M^t , $T_{b_i}^{a_j}$ is the transformation from point a_j to b_i , ω_{a_j} is the weight associated with mesh point a_j , which is the summation of each of its n_j member tracker weights:

$$\omega_{a_j} = \sum_{k=1}^{n_j} \omega_k \quad (5.12)$$

and ω_{tot} is the total weight of all mesh points a_j :

$$\omega_{tot} = \sum_{k=1}^m \omega_{a_k} \quad (5.13)$$

The second step is to compute the final target state S^t from the deformation mesh M^t by using the border points B . The target parameter vector $P = (p_1, \dots, p_5)$, is defined as follows:

- p_1 is the x coordinate of the target center.
- p_2 is the y coordinate of the target center.
- p_3 is the rotation about the center of the target.
- p_4 is the width of the target.
- p_5 is the height of the target.

The parameter vector P is computed from the border points B of the winning coalition C^t as follows:

$$p_1 = \frac{1}{4} \sum_{i=1}^4 b_{ix} \quad (5.14)$$

$$p_2 = \frac{1}{4} \sum_{i=1}^4 b_{iy} \quad (5.15)$$

$$p_3 = \frac{1}{|C|} \sum_{i=1}^{|C|} c_{i\theta} \quad (5.16)$$

$$p_4 = \frac{\sqrt{(b_{1x} - b_{2x})^2 + (b_{1y} - b_{2y})^2}}{2} + \frac{\sqrt{(b_{3x} - b_{4x})^2 + (b_{3y} - b_{4y})^2}}{2} \quad (5.17)$$

$$p_5 = \frac{\sqrt{(b_{1x} - b_{4x})^2 + (b_{1y} - b_{4y})^2}}{2} + \frac{\sqrt{(b_{2x} - b_{3x})^2 + (b_{2y} - b_{3y})^2}}{2} \quad (5.18)$$

where $|C|$ is the cardinality of the winning coalition.

5.2.4 Configuration of Tracking Network

By varying the number of trackers and their relative spatial location, one could produce a large number of possible configurations for the tracking network. We narrow down this to a manageable subset by considering only uniform grids over a rectangular region (see Fig. 5.6).

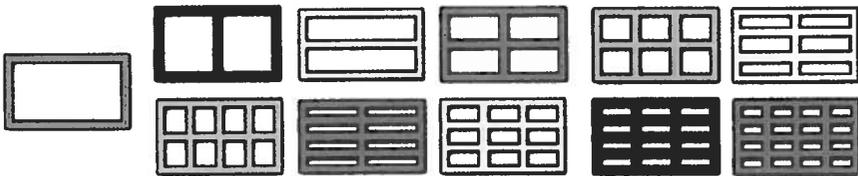


Fig. 5.6 Tracking network configurations selected for evaluation

An important consideration when picking a network configuration is the number of coalitions that must be evaluated to make a final target prediction because the number of coalitions increases with the number of trackers in the network (see Fig. 5.7).

To determine the relative performance of each of the configurations, we used each configuration to track the same target (i.e., a face) in a thermal video sequence. The true target position was annotated in each frame of the thermal video sequence to allow computation of the tracking errors from the various configurations (see Fig. 5.8).

The results show correlation between the number of trackers and tracking accuracy (see Fig. 5.9). Another trend in the data is that extra columns of trackers within the network configuration seem to be more beneficial than extra rows (see Fig. 5.9). This is explained by the facial motion exhibited in the particular experiment. The subject looks side to side and thereby deforms out of plane in the horizontal axis. Therefore, tracking configurations that add more detail along the horizontal dimension (i.e., more columns) perform better in the experiment.

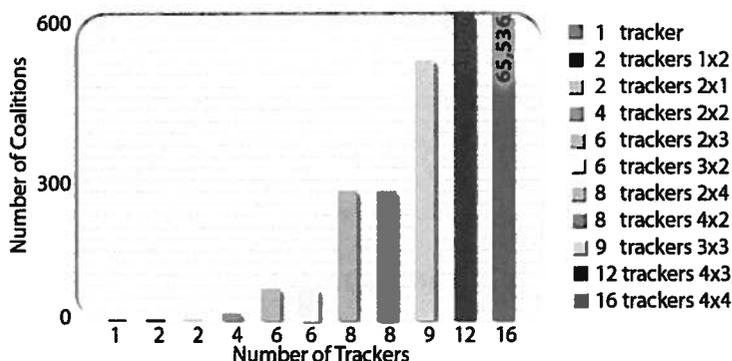


Fig. 5.7 Number of coalitions for each network configuration

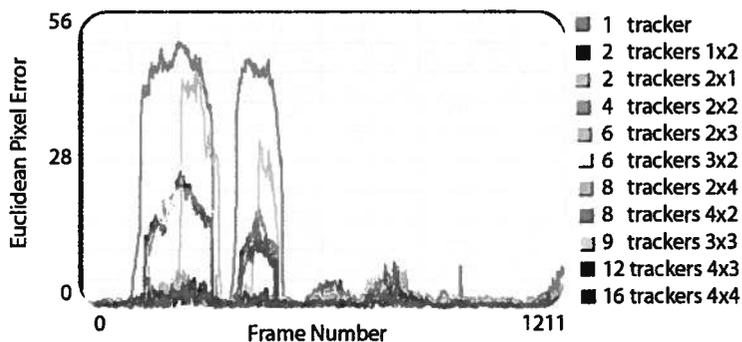


Fig. 5.8 Tracking errors of network configurations along the timeline; two large spikes correspond to out-of-plane movement by the target

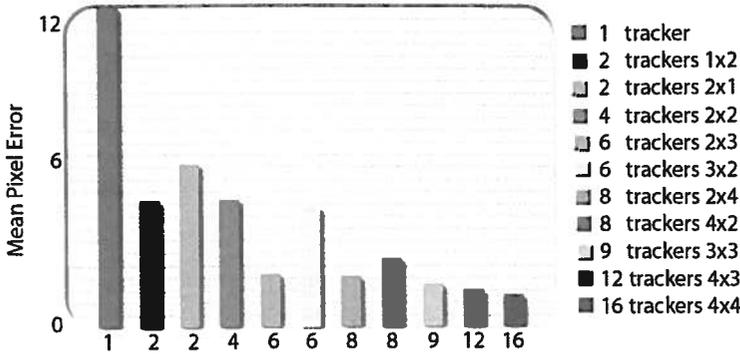


Fig. 5.9 Mean tracking error of network configurations as computed from the data in Fig. 5.8

Not surprisingly, there is not one tracking network configuration that fits every situation. The tracking network configuration must be selected based on the intended application and the type of target motion expected. In our case, we chose as a compromising solution a 3×3 tracking network configuration because we had a diverse assortment of target motions within our data set.

5.3 Experimental Design

An important consideration in our experimental design was exact quantification of the tracker's performance. For this, we needed an environment that would provide automatic ground-truthing. The cornerstone of our experimental design, however, was the provision to test our tracking method on video input from at least two different bands of the electromagnetic spectrum, one reflected and one radiated. The motivation was to demonstrate that the methodology is general enough to handle both. To satisfy this specification, we performed experiments using visual band video (reflected) and midwave infrared video (radiated). The underlying implication was that if the tracker worked on both visual and midwave infrared video, then the tracker would be general enough to be adapted to other radiated bands, such as long-wave infrared, as well as other reflected bands, such as the near infrared.

5.3.1 Design of Simulated Tracking Environment

We used a simulated tracking environment to precisely quantify the tracker's performance. The environment was initialized to a frame of thermal video, and then the tracker was initialized to the target. The target to be tracked was then translated about the image plane while simultaneously undergoing transformations. Because

the target transformations were dictated by the simulated environment, we could measure the true target state against its state projected by the tracker for each frame. Every simulated run was 200 frames in length.

5.3.2 Design of Thermal Infrared Experiment

For the purpose of testing the tracking algorithm on thermal infrared video, we selected a data set that was used in previous publications [12]. It consists of 39 video clips, each containing a main human subject undergoing an interview. We chose to track 1,000 frames of video from each of the subject clips, for a total of 39,000 frames of video. The chosen video segments featured a temporary occlusion of the main subject by another subject who was passing through the field of view. More important, the clips featured out-of-plane rotation of facial tissue as subjects were rotating their heads left or right, up or down. We chose a single-particle filter tracker to compare against the coalitional tracker. Both the single-particle filter and coalitional trackers featured identical parameterization. Both the single-particle filter and the coalitional network were tasked to track exactly the same facial tissue of each subject. The ground-truthing of this experiment was the reconciliation of the observations of two independent operators.

5.3.3 Design of Visual Experiment

To demonstrate that the tracking methodology can also be applied to visual band video, we performed experiments on a series of visual videos, each containing a different type of target. These targets included faces and cans.

5.4 Experimental Results

5.4.1 Results of Simulated Tracking Environment

We first measured the accuracy of the coalitional tracker (see Fig. 5.10). The coalitional tracker maintained a mean error of about 1 pixel, which is sufficient for demanding applications, such as physiological measurements [5]. Moreover, the coalitional tracker exhibited consistent performance over 20 identical trials (see Fig. 5.10). This is extremely important because it would be impossible to extract useful physiological measurements if the tracker gave inconsistent results each time it was run. However, due to the stochastic nature of particle filtering, it is very difficult to altogether eliminate minute variability from the tracking result. To determine

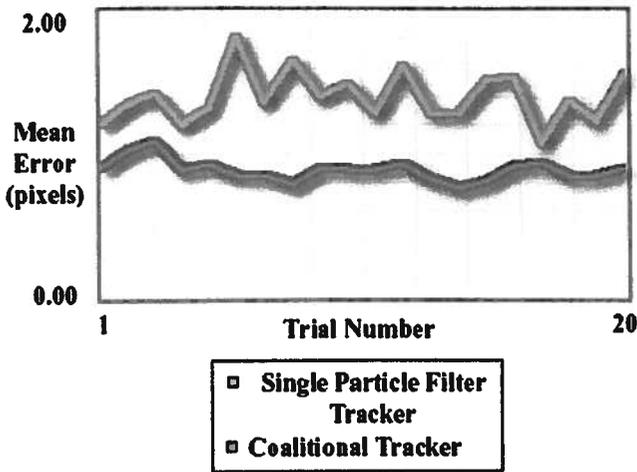


Fig. 5.10 Error and stability analysis of single-particle filter (green) versus coalitional tracking (red). Both trackers were used to track the same target in 20 identical trials using the simulated tracking environment

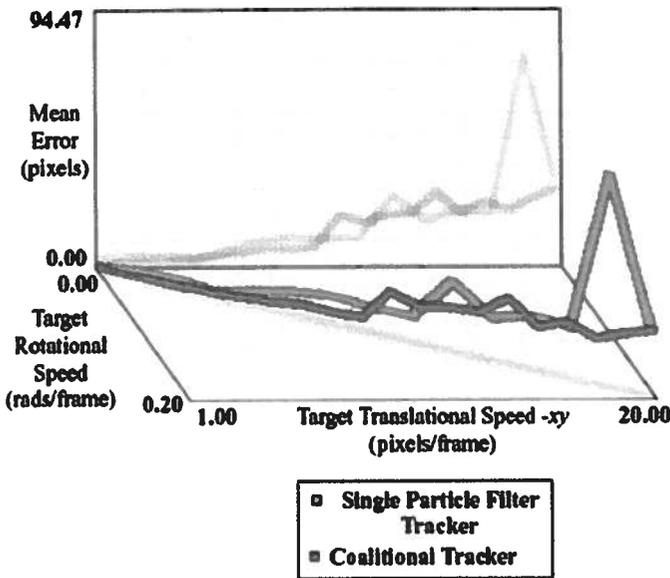


Fig. 5.11 Error analysis of single-particle filter (green) versus coalitional tracking using the tracking network (red). Both trackers were used to track the same target in 20 trials using the simulated tracking environment. Each trial involved increasingly faster translational and rotational target motion

the operational limits of the tracker, we measured its error under increasingly faster target motion in the simulated tracking environment (see Fig. 5.11). The superior performance of the coalitional tracker in complex and fast transformations is evi-

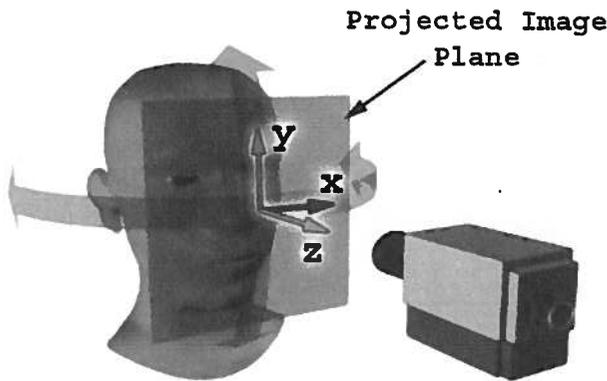


Fig. 5.12 Out-of-plane facial rotation. Any rotation that is not about the z-axis is considered out-of-plane rotation

dent. The coalitional tracker was also capable of negotiating out-of-plane facial rotations (see Fig. 5.12) much more successfully than the single-particle filter tracker (see Fig. 5.13).

5.4.2 Results of Thermal Infrared Experiment

The results from the thermal infrared experiment (see Fig. 5.14 and Table 5.1) clearly show that the coalitional tracker provides superior tracking over the single-particle filter tracker. The proposed method proved robust in typical (see Fig. 5.15) and difficult (see Fig. 5.16) operational scenarios. The few failures of the coalitional tracker were mainly caused by significant out-of-plane rotation or substantial occlusion of the target (see Fig. 5.17).

A rare case of failure is exemplified in Fig. 5.18, when the subject experienced rapid physiological changes on a grand scale. The subject in the figure underwent facial temperature increase in excess of 2°C within 6 min due to a state of high anxiety. This problem is due to the template measurement method, which assumes that the target's projection will not change dramatically over time. One possible solution to this problem is to dynamically update the template as presented in [39].

We extracted a sample physiological measurement from subject S2 (inventory reported in [12]) and compared it against the respective ground truth signal. The measured signal is the mean temperature of the subject's periorbital area through the course of the video clip. It is evident that the coalitional tracker enables the acquisition of a signal nearly identical to the ground truth (see Fig. 5.19), an indication of its fitness for accurate physiological measurements.

Fig. 5.13 Out-of-plane rotation comparison. *Left*, single-particle filter tracker (green); *right*, coalitional tracker (red). **a** Initial frame, **b** and **c** intermediate frames, **d** final frame in a 1-min thermal clip. The poor performance of the single-particle filter tracker is evident

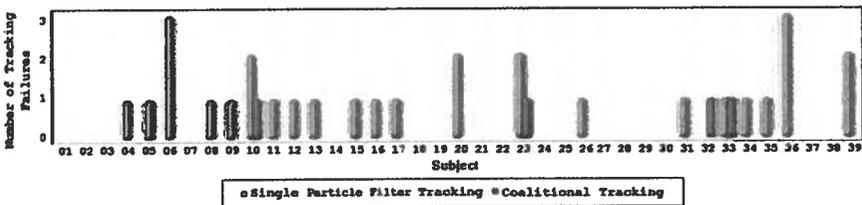
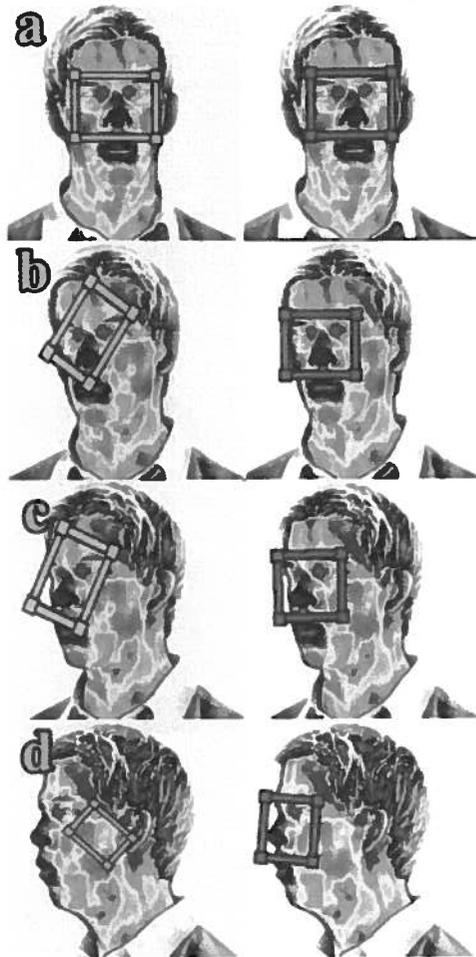
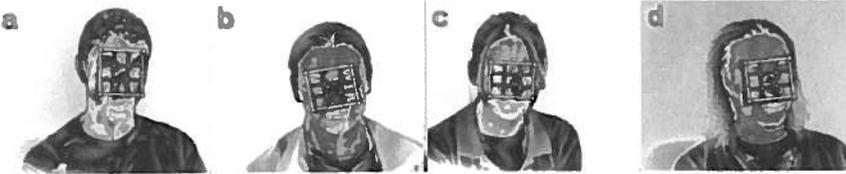
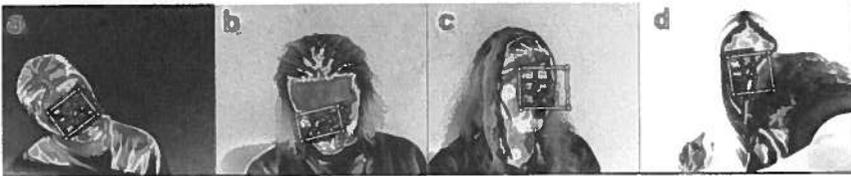


Fig. 5.14 Tracking failure graph for the 39 video clips in the thermal data set. For each clip the number of single-particle filter and coalitional tracking failures is shown in green and red, respectively. The absence of red bars in some video clip entries indicates perfect performance of the coalitional tracker

Table 5.1 Causation of tracking failures in the thermal data set

Reason for failure	Coalitional tracker failures	Single-tracker failures
Target rotation	1	18
Partial occlusion	2	9
No recovery	1	2
Total	4	29

**Fig. 5.15** Typical facial-tracking examples from the thermal data set. The selected subjects represent different ethnicities and both genders**Fig. 5.16** Successful coalitional tracking in the presence of difficult circumstances in the thermal spectrum: **a** target rotating in plane, **b** target rotating out of plane, **c** target rotating out of plane, **d** target partially occluded**Fig. 5.17** Tracking failures in the thermal spectrum. **a** and **c** The target has rotated out of plane beyond the tracker's ability to compensate. **b** The original target (periorbital area) is largely occluded. **d** The target has undergone extreme physiological changes relative to the initial tracking frame (see Fig. 5.18 for more details)

5.4.3 Results of Visual Experiment

The coalitional tracker performed robustly in several visual band experiments with various objects (faces and cans). The template was composed of 3-tuples (red, green, and blue reflectance values) instead of temperatures. The motion patterns included translation, rotation, and scaling (see Fig. 5.20 and Fig. 5.21).

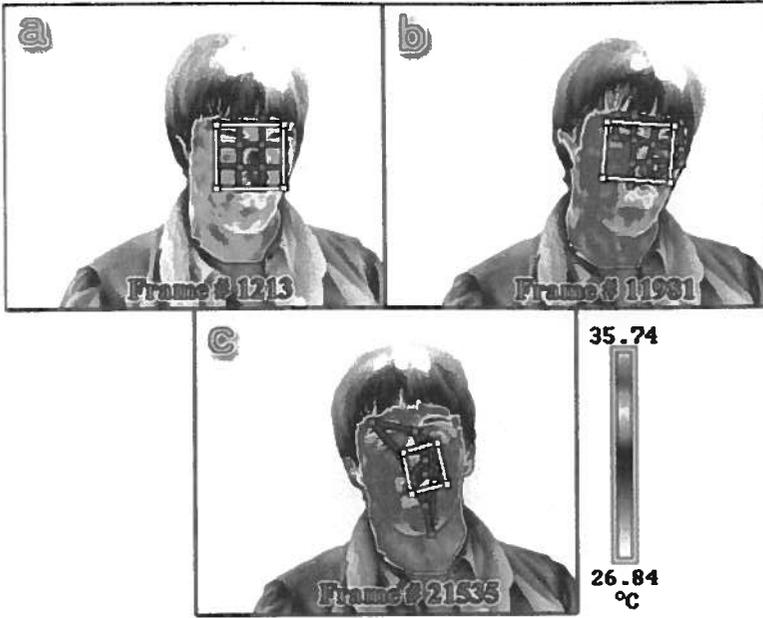


Fig. 5.18 Coalitional tracker performance under substantial physiological changes. **a** Tracker initialization. **b** The subject’s face undergoes a substantial thermal change in the middle of the video clip. The tracker is still performing correctly, but the winning coalition is composed of fewer trackers that are able to follow their targets. **c** Toward the end of the clip, the subject’s facial thermal profile continues to change dramatically, and the coalitional tracker is off target

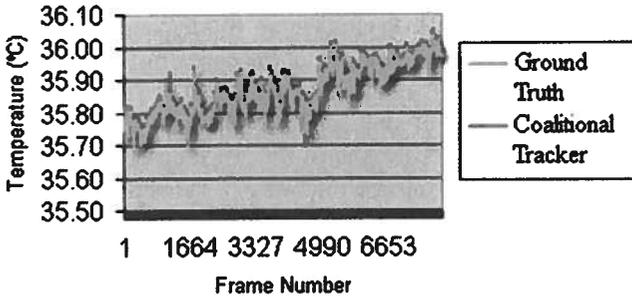


Fig. 5.19 Physiological signal extracted using coalitional tracking (in red) versus the ground truth signal

5.5 Application Perspective

Coalitional tracking is a fairly general framework and can be applied to a variety of problems across imaging modalities. Nevertheless, it was originally developed for a particular modality (i.e., thermal infrared) and for specific applications (i.e., physiological measurements on the face). To measure the success of coalitional tracking, it



Fig. 5.20 Example of tracking a face experiencing scaling and translation in the visual spectrum. The frames are shown chronologically from left to right. The deformation mesh is shown in blue, and the white rectangle represents the projected target state

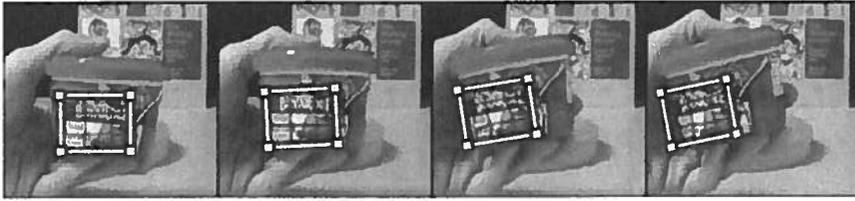


Fig. 5.21 Example of tracking an object experiencing scaling and translation in the visual spectrum. The frames are shown chronologically from left to right. The deformation mesh is shown in blue, and the white rectangle represents the projected target state

is important to understand its original application framework and impact. We touch on two major applications for which coalitional tracking is now used routinely: lie detection and sleep studies.

Levine et al. [8] reported a physiological sign of stress manifested as increased blood flow in the orbital muscle. Pavlidis et al. [9] demonstrated the potential of this stress sign as a lie detection indicator in the context of a well-designed interrogation. The importance of this cannot be overestimated. It was the first time that a localized physiological sign of cholinergic origin was identified on the face as “polygraph ready.” The facial locale is ideal for casual observation as it is typically exposed. Furthermore, the sympathetic relevance of the periorbital sign ranks very high because the face is heavily innervated with neuronal pathways. In summary, we had a primary stress indicator easily observable, but unfortunately not easily measurable. To begin, there was significant difficulty in sensing blood flow in the orbital muscle because of the delicate nature of the tissue. A popular method for sensing blood flow is ultrasound. Imagine, for example, the examiner rubbing the eyes of the subject with an ultrasound wand to get a blood flow measurement on the orbital muscle. This would clearly be impractical, particularly in the context of psychophysiological experiments. The problem was solved with the introduction of thermal imaging as the modality of choice for such measurements. Superficial blood flow under thin facial tissue emits a heat signature due to convection, which can be captured and analyzed by a thermal imaging sensor package. Thermal imaging did not only solve the sensing problem in a practical way, but also superbly supported psychophysiological experiments because it is totally unobtrusive.

The fact that the measurement was done at a distance might have been a blessing from the psychological point of view, but it posed a challenge from the medical point of view. What was needed was a virtual probe to isolate the area of interest in the image — a nontrivial segmentation problem. Moreover, a tracking method had to be developed to keep this virtual probe in the orbital area irrespective of head motion. Both segmentation and tracking had to maintain pixel-level accuracy for the measurement to remain valid. The tracking problem, which is of interest here, was especially challenging due to the functional nature of thermal infrared imaging and the real-time requirements of the application. Thermal imaging of the face depicts physiological changes. Therefore, it is highly dynamic, unpredictable, and difficult to model. Despite the absence of strong models, tracking still has to be accurate enough to support valid medical measurements. Plus, it has to be highly efficient as the technology created an opportunity for lie detection “on the fly,” which the polygraph community wanted to fully exploit.

Coalitional tracking solved these conflicting requirements and facilitated research and development in a big way. The secret of its success is that it efficiently optimized the behavior of many weak model trackers to achieve robustness and accuracy reminiscent of strong model trackers in structural imaging domains (e.g., computed tomography). Coalitional tracking was used to measure stress in three major government experiments involving multiple lie detection interviews of more than 150 subjects. This accounts for hundreds of recording hours and millions of frames. All the thermal videos used in the experiments detailed in Section 5.4.2 are a small subset of the lie detection inventories.

Starting in February 2007, coalitional tracking was also applied with great success in sleep studies at the University of Texas Medical School. The physiological measurement of interest in this case was breathing (see Fig. 5.22). Patients in sleep studies suffer from chronic respiratory diseases that manifest themselves during sleep. A prime example of such a disease is obstructive sleep apnea, for which breathing is suspended for a few seconds, several times every minute. This creates temporary asphyxiation, which triggers the “fight-or-flight” response. As the phenomenon repeats itself every few seconds, it results in an almost permanent load

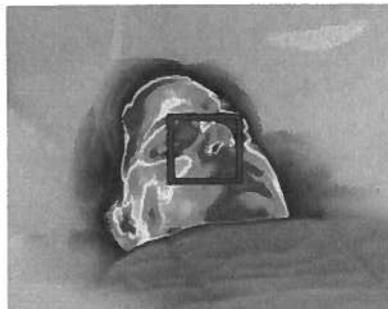


Fig. 5.22 Coalitional tracker monitoring a subject’s nasal region during inventory of sleep study experiments at the University of Texas Medical School

on the cardiovascular system and poor-quality sleep. Both have serious long-term repercussions on the health of the patient. Diagnosis of sleep apnea involves monitoring of the patient's sleep for several nights in the lab. During these times, the patient is heavily instrumented (see Fig. 5.23), a highly uncomfortable proposition for anyone, but especially for people who suffer from sleep problems. Therefore, there is strong motivation to unwire the patient to the extent possible.

Coalitional tracking helped to reliably extract the breathing signal through thermal imagery. The monitoring periods exceeded one h for every patient. The accuracy of the imaging computation was ascertained against the clinical gold standard (i.e., thermistor). In contrast to the lie detection application in which the subject's head moves in moderate amounts all the time, during sleep studies the patient's head exhibits minute motor motion (i.e., due to breathing) and occasionally abrupt large-scale motion (i.e., turning). The different motion profiles in the two applications represent a comprehensive testing of the tracker's abilities.



Fig. 5.23 Patient wired for sleep study (Courtesy University of Texas Medical School)

5.6 Conclusion

We have proposed a novel tracking method. Our method uses a spatially distributed network of trackers whose interactions are modeled using coalitional game theory. The output of the method provides pixel-level tracking accuracy, even in the presence of multidimensional target transformation.

We tested our method in thermal and visual video sets featuring faces and objects. We compared the performance of the proposed coalitional tracker with that of a single-particle filter tracker. The coalitional tracker exhibited superior performance in both regular and challenging tasks. The strength of the method comes from the redundancy that is elegantly encoded in its game theoretic structure. Detailed quantification and ground truth verification indicated that the new method provides accuracy appropriate for demanding medical imaging applications. Equally important is the fact that the method appears to be general and flexible enough to use in imaging applications across the electromagnetic spectrum.

5.6.1 Future Work

The particular adaptation of game theory to tracking presented in this chapter is but one of many possible approaches that might be adopted. For example, the problem of tracking could be alternatively viewed as a noncooperative game in which the trackers compete with each other, and the final solution could then be modeled as a Nash (strategic) equilibrium [29]. Plus, active research areas in game theory, such as stochastic and differential games [41], could potentially be adapted for use in tracking.

An important area that is amenable to improvement is the current static template scheme. Although it works well in the thermal infrared band, where emission of most objects does not change dramatically in short observation periods (e.g., a few minutes), it is potentially vulnerable in the visual band, where reflected light may change dramatically over a split second depending on the angle of incidence. A dynamic template mechanism will eliminate this vulnerability.

The current method is based on deterministic management of probabilistic trackers. A future method may be developed that will be based on probabilistic management of probabilistic trackers. This can be realized within a Bayesian framework in which the posterior weight of each tracker in the coalitional game would be computed from its prior and an appropriate likelihood function. Since this will add probabilistic memory into coalition membership, one can eliminate the membership retention factor in the current characteristic function, which in essence crudely plays the same role.

Acknowledgments We would like to thank the National Science Foundation (grant IIS-0414754) and Dr. Ephraim Glinert for supporting this effort in its early phase. We would also like to thank Dr. Andrew Ryan, Dr. Dean Pollina, Dr. Troy Brown, and the Department of Defense (multiple

research contracts) for supporting the late phase of the research. The views expressed by the authors in this chapter do not necessarily reflect the views of the funding agencies.

Chapter's References

1. D.A. Gonzalez, F.J. Madruga, M.A. Quintela, and J.M. Lopez-Higuera, Defect assessment on radial heaters using infrared thermography, *NDT & E International*, 38(6):428-432, September 2005
2. M. Burrell, Computer vision for high-speed, high-volume manufacturing, in *Proceedings of the 1993 International Conference on Systems, Man, and Cybernetics*, 3:349-354, October 17-20, 1993
3. I. Pavlidis, V. Morellas, P. Tsiamyrtzis, and S. Harp, Urban surveillance systems: from the laboratory to the commercial world, *Proceedings of the IEEE*, 89(10):1478-1497, October 2001
4. R.T. Collins, A.J. Lipton, H. Fujiyoshi, and T. Kanade, Algorithms for cooperative multi-sensor surveillance, *Proceedings of the IEEE*, 89(10):1456-1477, October 2001
5. M. Garbey, A. Merla, and I. Pavlidis, Estimation of blood flow speed and vessel location from thermal video, in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:356-63, June 27-July 2, 2004
6. N. Sun, M. Garbey, A. Merla, and I. Pavlidis, Imaging the cardiovascular pulse, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:416-21, June 20-25, 2005
7. J. Fei, Z. Zhu, and I. Pavlidis, Imaging breathing rate in the CO₂ absorption band, in *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, September 1-4, 2005
8. J. Levine, I. Pavlidis, and M. Cooper, The face of fear, *Lancet*, 357(9270), June 2, 2001
9. I. Pavlidis, N.L. Eberhardt, and J. Levine, Human behavior: seeing through the face of deception, *Nature*, 415(6867):35, January 3, 2002
10. I. Pavlidis and J. Levine, Thermal image analysis for polygraph testing, *IEEE Engineering in Medicine and Biology Magazine*, 21(6):56-64, November-December 2002
11. C. Eveland, D. Socolinsky, and L. Wolff, Tracking human faces in infrared video, *Image and Vision Computing*, 21:578-590, July 2003
12. P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, M.G. Frank, and P. Ekman, Lie detection—recovery of the periorbital signal through tandem tracking and noise suppression in thermal facial video, in *Proceedings of SPIE Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IV*, E.M. Carapezza, editor, p. 5778, March 29-31, 2005
13. S. Krotosky, S. Cheng, and M. Trivedi, Face detection and head tracking using stereo and thermal infrared cameras for "smart" airbags: a comparative analysis, in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems*, 1:17-22, 2004
14. A. Doucet, N. DeFreitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2001
15. M. Isard and A. Blake, Condensation — conditional density propagation for visual tracking, *International Journal of Computer Vision*, 19(1):5-28, 1998
16. M. Isard and A. Blake, ICONDENSATION: unifying low-level and high-level tracking in a stochastic framework, in *Proceedings of the 5th European Conference on Computer Vision*, 1:893-908, June 2-6, 1998
17. J. MacCormick and M. Isard, Partitioned sampling, articulated objects, and interface-quality hand tracking, in *Proceedings of the 7th European Conference on Computer Vision*, 1843:3-19, 2000

18. Y. Zhong, A.K. Jain, and M.P. Dubuisson-Jolly, Object tracking using deformable templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):544–549, May 2000
19. Y. Shi and W. Karl, Real-time tracking using level sets, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:34–41, June 20–25, 2005
20. C. Zimmer and J. C. Olivo-Marin, Analyzing and capturing articulated hand motion in image sequences, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11):1838–1842, November 2005
21. S. Goldenstein, C. Vogler, J. Stolfi, V. Pavlovic, D. Metaxas, Outlier rejection in deformable model tracking, in *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition*, June 19–26, 2004
22. T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, June 2001
23. F. Dornaika and J. Ahlberg, Efficient active appearance model for real-time head and facial feature tracking, *Proceedings of the 2003 IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 173–180, October 13, 2003
24. C. Cheng, R. Ansari, and A. Khokhar, Multiple object tracking with kernel particle filter, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:566–573, June 20–25, 2005
25. Y. Ting and W. Ying, Decentralized multiple target tracking using netted collaborative autonomous trackers, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:939–946, June 20–25, 2005
26. M. Isard and J. MacCormick, BraMBLe: a Bayesian multiple-blob tracker, in *Proceedings of the 8th IEEE International Conference on Computer Vision*, 2:34–41, July 7–14, 2001
27. J. MacCormick and A. Blake, A probabilistic exclusion principle for tracking multiple objects, *International Journal of Computer Vision*, 39(1):57–71, 2000
28. T.S. Ferguson, game theory, Chapter 4, http://www.math.ucla.edu/~tom/Game_Theory/Contents.html
29. K. Ritzberger, *Foundations of Non-Cooperative Game Theory*, Oxford University Press, New York, 2002
30. A. Rapoport, *N-Person Game Theory: Concepts and Applications*, University of Michigan, 1978
31. E. Rasmusen, *Games and Information: An Introduction to Game Theory*, Blackwell, 1989
32. T.G. Fisher et al., *Managerial Economics: A Game Theoretic Approach*, Routledge, 2002
33. C. Schmidt, editor, *Game Theory and Economic Analysis: A Quiet Revolution in Economics*, Routledge, 2002
34. P. Ordeshook, *Game Theory and Political Theory: An Introduction*, Cambridge University Press, Cambridge, U.K., 1986
35. S. Brams, *Game Theory and Politics*, Free Press, New York, 1975
36. S. Hart, editor, *Cooperation: Game-Theoretic Approaches*, Springer-Verlag, New York, 1997
37. M. Mareš, *Fuzzy Cooperative Games*, Physica-Verlag, 2001
38. S. Baker and I. Matthews. Equivalence and efficiency of image alignment algorithms, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1:1090–1097, 2001
39. I. Matthews, T. Ishikawa, S. Baker. The template update problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):810–815, June 2004
40. Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 19(7):721–732, 1997
41. M. Bardi, T. Raghavan, T. Parthasarathy, editors, *Stochastic and Differential Games: Theory and Numerical Methods*. *Annals of the International Society of Dynamic Games*, Birkhauser, 1998