Multi-Object Tracking via Species Based Particle Swarm Optimization

Xiaoqin Zhang[1], Weiming Hu[1], Wei Li[1], Wei Qu[2], Steve Maybank[3]
[1]National Laboratory of Pattern Recognition, Institute of Automation, Beijing, China
{xqzhang,wmhu,weili}@nlpr.ia.ac.cn
[2]Siemens Medical Solutions USA Inc., Illinois, USA
weiqu@siemens.com
[3]School of Computer Science and Information Systems, Birkbeck College, London, UK
sjmaybank@dcs.bbk.ac.uk

Abstract

Multiple object tracking is particularly challenging when many objects with similar appearances occlude one another. Most existing approaches concatenate the states of different objects, view the multi-object tracking as a joint motion estimation problem and search for the best state of the joint motion in a rather high dimensional space. However, this centralized framework suffers a great computational load. We brings a new view to the tracking problem from a swarm intelligence perspective. In analogy with the foraging behavior of the bird flocks, we propose a species based PSO (particle swarm optimization) algorithm for multiple object tracking, in which the global swarm is divided into many species according to the number of objects, and each species searches for its object and maintains it. The interaction between different objects is modeled as species competition and repulsion, and the occlusion relationship is implicitly deduced from the 'power' of each species, which is effectively evaluated by the image observations. Therefore, our approach decentralizes the joint tracker to a set of individual trackers, each of which tries to maximize its visual evidence. Experimental results demonstrate the efficiency and effectiveness of our method.

1. Introduction

Multiple object tracking in videos is one of the most important problems in many emerging applications, such as surveillance, intelligent transportation, human-computer interface, and video analysis. Due to its crucial value in these applications, many efforts have been made to solve this problem in the last decades [1-17]. MacCormick and Blake [3] develop a probabilistic exclusion principle to solve the association problem in multiple object tracking, but it can only be applied to pairs of objects. BraMBLe [4], a Bayesian multi-blob tracker combines a multi-blob likelihood function with a particle filter. Khan et al. [6] propose an MCMC-based (Markov chain Monte Carlo) particle filter which uses a Markov random field to model motion interaction, but their model can not deal with occlusion. Yu et al. [7] propose a collaborative tracking algorithm for multiple objects which models the joint prior of objects using a Markov random network to solve the identity maintenance problem. Qu et al. [9] suggest an interactively distributed multi-object tracking algorithm using a magnetic-inertia potential model to solve the multiple object labeling problem in the presence of occlusions. In [10], the spatio-temporal context of each object is used to maintain the correct identification of the object. Nillius et al. [11] propose a method to resolve multiple hypotheses via Bayesian networks and a novel solution is obtained by belief propagation techniques. Yang et al. [12] propose a game-theoretic multiple target tracking algorithm, in which the tracking problem is solved by finding the Nash Equilibrium of a game. It can decentralize the joint tracker and uses computational resources efficiently. Another solution to overcome the curse of dimensionality in tracking multiple objects jointly is the variational particle filter proposed by Jin et al. [13], where the proposal distribution is based on the approximated posterior obtained by variational inference. Despite the increasing amount of work done in this field, multiple object tracking still remains challenging for the following reasons: 1) it is difficult to handle the interactions and the occlusions among different objects; 2) it is difficult to reacquire a severely occluded object after the occlusion process; 3) nuisance factors, such as low quality video and camera parameter changes also bring extra difficulties.

Recently PSO (particle swarm optimization) [18, 19], a new population based stochastic optimization technique, has received more and more attention because of its considerable success. Unlike the independent particles in the particle filter, the particles in PSO interact locally with one

2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops
978-1-4244-4441-0/09/$25.00 ©2009 IEEE

another and with their environment in analogy with the cooperative and social aspects of animal populations, for example as found in birds flocking. Starting from a diffuse population, now called a swarm, individuals, now termed particles, tend to move in the state space and eventually cluster in regions where optimal state is located. The advantages of this mechanism are, on one hand, the robustness and sophistication of the obtained group behavior and, on the other hand, the simplicity and low cost of the computation associated with each particle.

Inspired by the forgoing discussions, we propose a species based PSO algorithm for multiple object tracking, where the global particle swarm is divided into several species according to the number of objects. The main contributions of the proposed tracking algorithm are summarized as follows:

1. A species concept is introduced into the PSO framework to extend it to multiple object tracking. The particles are divided into species such that each species corresponds to one of the objects. The occlusion between different objects is modeled as species competition, and the occlusion relationship is implicitly deduced from the power of each species. Meanwhile, a repulsion force is employed to prevent the particles in one species from being miss-attracted by other species. As a result, the joint tracker can be decentralized to the individual trackers, which try to maximize their visual evidence.

2. The appearances of objects under occlusion are carefully updated based on the corresponding observation models. Thus an object emerging from severe occlusion can be successfully reacquired after occlusion.

This paper is organized as follows. The motivation of our approach is given in Section 2. A brief introduction to the traditional PSO algorithm is presented in Section 3. In section 4, we show the details of our proposed tracking approach. Experimental results are shown in Section 5, and Section 6 is devoted to conclusion.

2. Motivation

2.1. Single Object Tracking From the Biology Swarm Intelligence Viewpoint

First, we define the following analogies: 1) the groundtruth state of an object and its support region are viewed as ecological resources (e.g. food); 2) the particles in state space correspond to a certain animal (e.g. bird); 3) the observation likelihood of each particle is analogous to fitness ability of an individual animal. Then the single object tracking problem is viewed in the following way: suppose a group of particles (birds) are randomly generated in the image (state space), and none of the particles (birds) knows where the object (food) is. But each particle (bird) knows how far it is from the object (food) by evaluating the observation (fitness ability) in each iteration. What is the best strategy to find the object (food), and how can the information obtained by each particle (bird) be used efficiently? The PSO framework, inspired by the swarm intelligence of birds flocking, provides an effective way to answer these questions.

2.2. Extended to Multiple Objects Tracking

When the multiple objects are separated, the mechanism in Section 2.1 for single object tracking can be easily extended to multiple object tracking by creating a tracker for each object, and these trackers are conducted independently. If the objects move close together and even occlude each other, the independent sequential PSO framework may fail. As mentioned in Section 2.1, the support regions of objects are analogous to ecological resources, e.g. food. If occlusion happens between two objects, their support regions overlap, which means that the overlap part is the resource needed by both species. Consequently, the competition and repulsion between the species arises as they compete for this part of the resource, and the stronger species may win this competition with a higher probability. From the discussions of the relationship between multiple object tracking and biological swarm, we can find that our assumption and analogies are reasonable and tractable.

In the following two sections, we first briefly review the traditional PSO algorithm, and then give a detailed description of the multiple species based particle swarm optimization tracking algorithm.

3. Particle Swarm Optimization

Particle swarm optimization [18], is a population based stochastic optimization technique, which is inspired by the social behavior of bird flocking. In detail, a PSO algorithm is initialized with a group of random particles $\{x^{i,0}\}_{i=1}^{N}$ (N is the number of particles). Each particle $x^{i,0}$ has a corresponding fitness value $f(x^{i,0})$ and has a relevant velocity $v^{i,0}$, which is a function of the best state found by that particle (p^i, for individual best), and of the best state found so far among all particles (g, for global best). Given these two best values, the particle updates its velocity and state with following equations in the nth iteration (as shown in Fig.1),

$$v^{i,n} = \mathcal{R}(v^{i,n-1} + \varphi_1 u_1 (p^i - x^{i,n-1}) + \varphi_2 u_2 (g - x^{i,n-1})) \quad (1)$$

$$x^{i,n} = x^{i,n-1} + v^{i,n} \quad (2)$$

where the φ_1, φ_2 are acceleration constants, $u_1, u_2 \in (0, 1)$ are uniformly distributed random numbers, and $\mathcal{R}$ is a constriction factor to confine the velocity within a reasonable range: $||v^{i,n}|| \leq v^{max}$. After evaluating the fitness value of

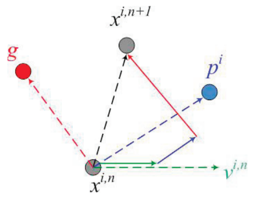

Figure 1. The nth iteration of particle i

each particle, the individual best and global best of particles are updated in the following equations:

$$p^i = \begin{cases} x^{i,n}, & \textbf{if } f(x^{i,n}) > f(p^i) \\ p^i, & \textbf{else} \end{cases}, \quad g = \arg\max_{p^i} \ f(p^i) \tag{3}$$

In Eq.(1), the three different parts represent *inertial velocity*, *cognitive effect* and *social effect* respectively. As a result, the particles interact locally with one another and with their environment in analogy with the 'cognitive' and 'social' aspects of animal populations, and eventually cluster in the regions where the optima are located.

4. Proposed Tracking Approach

In our tracking algorithm, the motion of a tracked object between two consecutive frames is approximated by a set of affine parameters $(t_x, t_y, \theta, s, \alpha, \beta)$, where $\{t_x, t_y\}$ denote the 2-D translation parameters and $\{\theta, s, \alpha, \beta\}$ are deformation parameters. A particle is a sample from the affine parameter space and its fitness value is evaluated by a subspace based appearance model [20]. In the following parts, we first introduce the incremental subspace learning based appearance model, and then give a detailed description of the proposed multi-object tracking algorithm.

4.1. Incremental Subspace Learning Based Appearance Model

In this part, we introduce a subspace based appearance model [20] for observation evaluation, which models the appearance of an object by incrementally learning a low-order eigenspace representation.

Observation Likelihood: As shown in [20], given the learned subspace U and a new observation o_t, the observation likelihood is based on the reconstruction error of the observation o_t in the object subspace, which is defined as follows.

$$RE = ||o_t - UU^T o_t||^2 \tag{4}$$

As a result, the observation likelihood is naturally formed as

$$p(o_t|x_t) = exp(-RE) \tag{5}$$

Incremental Subspace Learning: Given the SVD of the previous appearance data $A = \{o_1, \dots, o_t\}$, i.e. $A = U\Sigma V^T$, where each column o_i is the observation of the object in the *ith* frame. After tracking k frames, we have obtained k most recent observations of the object $E = \{o_{t+1}, \dots, o_{t+k}\}$, the R-SVD algorithm [21] efficiently

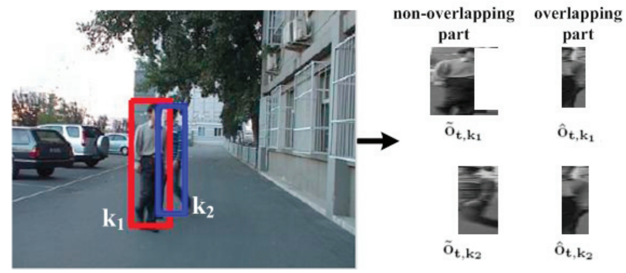

Figure 2. Overlapping region of two objects under occlusion.

computes the SVD of the matrix $A^{'} = (A|E) = U^{'}\Sigma^{'}V^{'T}$ based on the SVD of A as follows:

1. Apply QR decomposition to E and get orthonormal basis $\tilde{E}$, and $U^{'} = (U|\tilde{E})$.
2. Let $V^{'} = \begin{pmatrix} V & 0 \\ 0 & I_k \end{pmatrix}$ where I_k is a $k \times k$ identity matrix. It follows then,

$$\Sigma^{'} = U^{'T}A^{'}V^{'} = \begin{pmatrix} U^T \\ \tilde{E} \end{pmatrix} (A|E) \begin{pmatrix} V & 0 \\ 0 & I_k \end{pmatrix}$$

$$= \begin{pmatrix} U^T AV & U^T E \\ \tilde{E}^T AV & \tilde{E}^T E \end{pmatrix} = \begin{pmatrix} \Sigma & U^T E \\ 0 & \tilde{E}^T E \end{pmatrix}$$

3. Compute the SVD of $\Sigma^{'} = \tilde{U}\tilde{\Sigma}\tilde{V}^T$ and the SVD of $A^{'}$ is

$$A^{'} = U^{'}(\tilde{U}\tilde{\Sigma}\tilde{V}^T)V^{'T} = (U^{'}\tilde{U})\tilde{\Sigma}(\tilde{V}^T V^{'T})$$

In this way, the R-SVD algorithm computes the new eigenbasis efficiently.

4.2. Multiple Object Tracking Algorithm

As stated in Section 2.2, in the multi-object tracking case, the observations of different objects may overlap during occlusion, and the correspondences between objects and their features become ambiguous. To overcome this difficulty, we propose a multiple species based PSO algorithm. The fundamental idea of the proposed algorithm is to divide the particles into several species according to the number of objects, and effectively model the interactions and the occlusions between different species.

Below we give a detailed description of our algorithm which contains the following parts: 1) problem formulation; 2) competition and repulsion model; 3) annealed Gaussian based PSO; 4) selective updating for the appearance model.

4.2.1 Problem Formulation

Let's recall the symbols for the states and observations $\mathcal{X} = \{x^{i,n}_{t,k}, i = 1, \dots, N, k = 1, \dots, M\}, \mathcal{O} = \{o^{i,n}_{t,k}, i = 1, \dots, N, k = 1, \dots, M\}$, where N is the number of particles and M is the number of objects. Thus multiple object tracking can be formulated as follows.

$$\mathcal{X}^* = \arg\max_{\mathcal{X}} p(\mathcal{O}|\mathcal{X}) \tag{6}$$

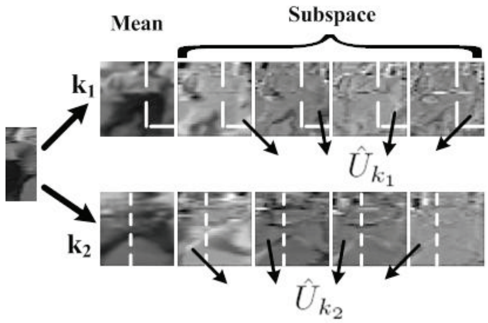

Figure 3. Project the overlapping part to the corresponding part of learned subspaces.

If no occlusion happens, the above optimization problem can be simplified by maximizing the individual observation likelihood independently (here, we drop the superscript i, n for simplicity):

$$x^*_{t,k} = \arg\max_{x_{t,k}} p(o_{t,k}|x_{t,k}), k = 1, \ldots, M \tag{7}$$

If occlusion happens, then we divide the observation of object under occlusion into two parts: 1) non-overlapping part $\tilde{o}_{t,k}$; 2) overlapping part $\hat{o}_{t,k}$. For example, the occlusion happens between objects k_1 and k_2 (see Fig.2), thus the tracking problem of these two objects can be formulated as follows:

$$x^*_{t,k_1} = \arg\max_{x_{t,k_1}} p(\tilde{o}_{t,k_1}|x_{t,k_1})p(\hat{o}_{t,k_1}|x_{t,k_1}, x_{t,k_2}) \tag{8}$$

$$x^*_{t,k_2} = \arg\max_{x_{t,k_2}} p(\tilde{o}_{t,k_2}|x_{t,k_2})p(\hat{o}_{t,k_2}|x_{t,k_2}, x_{t,k_1}) \tag{9}$$

where $p(\hat{o}_{t,k_1}|x_{t,k_1}, x_{t,k_2})$ and $p(\hat{o}_{t,k_2}|x_{t,k_2}, x_{t,k_1})$ are the interactive likelihood of the corresponding object on the overlapping part respectively. The Eq.(8) and Eq.(9) are iteratively computed until convergence[1].

4.2.2 Competition and Repulsion Model

Competition Model: When occlusion between different objects happens, the corresponding support regions may overlap (see Fig.2). In this case, the competition between two species arises to scrabble for the overlapping part. The question is how to effectively model the competition phenomenon in tracking problems?

In order to answer the above question, we first need to tie the visual problem to this phenomenon, and model the detail of the competition process. Before introducing our model, we first discuss the two related works [9, 12]. Although they both model the interactions among the objects through the observations, the detailed models are totally different from ours. In [9], the interaction is modelled using the whole support of the observation region, not just the overlapping part. This is a little unreasonable, since the competition only happens in the overlapping region, and the effectiveness of the model may be diluted by the non-overlapping regions. In contrary, [12] models the interference only using the overlapping part. However, the model focuses on the pixel level. It may not be very robust when the interacting objects have a similar color or are under severe occlusion. In this paper, we view the overlapping part as a whole and project it onto the corresponding part of the learned subspace of each object (see Fig.3), and the fitness value on the overlapping part is evaluated as the competition ability. As a result, the power of each species is defined as follows

$$power^k = p(\hat{o}_{t,k}|x_{t,k}) = exp(-||\hat{o}_{t,k} - \hat{U}_k\hat{U}_k^T\hat{o}_{t,k}||^2) \tag{10}$$

where $\hat{U}_k$ is the the corresponding subspace of the overlapping part of the object k. Consequently, the interactive likelihood $p(\hat{o}_{t,k_1}|x_{t,k_1}, x_{t,k_2})$ of object k_1 on the overlapping parts can be obtained[2]:

$$p(\hat{o}_{t,k_1}|x_{t,k_1}, x_{t,k_2}) = \frac{power^{k_1}}{\sum_{i=1,2} power^{k_i}} \tag{11}$$

The competition ability can be described by the interactive likelihood for each species. A species with higher competition ability is more likely to win the competition, which means that the object corresponding to this species is more likely to be the one occluding other objects. We will validate this conclusion through the experiments (see Section 5.2.2).

Species Repulsion: Generally, multiple object tracking algorithms suffer from the the well-known coalescence problem during occlusion, where a tracker loses its associated object and falsely tracks other objects. While in the real world, the stronger species will repel other species and try to take up all the resources. In order to tackle the coalescence problem, we need to define a repulsion model for the objects under occlusion. When occlusion happens between objects k_1 and k_2, the repulsion force from object k_2 to object k_1 is defined as follows.

$$F_{\overrightarrow{k_2,k_1}} = p(\hat{o}_{t,k_2}|x_{t,k_1}, x_{t,k_2})V_{k_1} \tag{12}$$

where V_{k_1} is a velocity vector of object k_1. The scale parameter $p(\hat{o}_{t,k_2}|x_{t,k_1}, x_{t,k_2})$ is determined by the competition ability of object k_2, representing the intension of the repulsion force. The species repulsion model means that the species with a larger the competition ability repels other species nearby with a larger force.

This repulsion force is added to the particle evolution process (see Section 4.2.3) to prevent the particle from being miss-attracted by other species and maintain the diversity among the species. In this way, the competition model

[1] The occlusion between three or more objects can be formulated similarly.

[2] Here, we also assume that the occlusion happens between object k_1 and object k_2.

is incorporated into the particle evolution process, thus alleviating the coalescence problem.

4.2.3 Annealed Gaussian Based PSO

Since the conventional PSO contains many parameters which need to be tuned carefully, we adopt an annealed Gaussian based particle swarm optimization (AGPSO) algorithm [22], where the particles and their velocities are updated in the following way,

$$v^{i,n+1} = |r_1|(p^i - x^{i,n}) + |r_2|(g - x^{i,n}) + \epsilon \tag{13}$$

$$x^{i,n+1} = x^{i,n} + v^{i,n+1} \tag{14}$$

where $|r_1|$ and $|r_2|$ are the absolute values of the independent samples from the Gaussian probability distribution $\mathcal{N}(0,1)$, and ϵ is zero-mean Gaussian perturbation noise to avoid trapping in local optima. The covariance matrix of ϵ is changed in an adaptive simulated annealing way [23]:

$$\Sigma_\epsilon = \Sigma e^{-cn} \tag{15}$$

where Σ is the covariance matrix of the predefined transition distribution, c is an annealing constant, and n is the iteration number. The elements in Σ_ϵ decrease according to the iteration number which enables a fast convergence rate.

If occlusion happens between object k_1 and k_2 at time t, we add a repulsion force to the particle evolution process, and then the iteration form for object k_1 becomes as follows:

$$v^{i,n+1}_{t,k_1} = |r_1|(p^i_{t,k_1} - x^{i,n}_{t,k_1}) + |r_2|(g_{t,k_1} - x^{i,n}_{t,k_1}) + |r_3| F_{\overrightarrow{k_2,k_1}} + \epsilon$$

$$x^{i,n+1}_{t,k_1} = x^{i,n}_{t,k_1} + v^{i,n+1}_{t,k_1}$$

where r_3 is also the Gaussian random number sampled from $\mathcal{N}(0,1)$. The third part on the right hand side of the above equation represents the interactive effect from object k_2 to object k_1.

In summary, our approach models the competition phenomenon on the observation level and models competition effect on the state space to guide the evolution process of object state. Experiments show that our model is reasonable.

4.2.4 Selective Updating for Appearance Model

In most multiple tracking algorithms, updating of the appearance model is stopped during occlusion. However, if the object appearance changes during occlusion, the tracker may fail to reacquire the object after the occlusion. In our paper, we design a selective updating scheme to accommodate the appearance changes during occlusion: 1) as shown in the Fig.2, pixels in the visual part of objects are incrementally updated in the normal way, 2) pixels in the overlapping part are projected back to the corresponding subspace of each object (see Fig.3) and the reconstruction errors are calculated.

$$R = |\hat{o}_{t,k} - \hat{U}_k \hat{U}_k^T \hat{o}_{t,k}| \tag{16}$$

If the reconstruction error of a pixel inside the overlapping part is smaller than a predefined threshold, then it is updated to the corresponding subspace.

4.3. Algorithm Summary

Our proposed tracking algorithm is summarized as follows [3].

1. Initialization: $t = 0$, the states of the multiple objects are manually initialized as the global best for species $\{g_{t,k}\}_{k=1}^M$. The individual best $\{p^i_{t,k}\}_{i=1}^N$ are set equal to $g_{t,k}$.
2. **for** $t = 1, 2, \ldots$ **do**
3. Check occlusions among $\{g_{t-1,k}\}_{k=1}^M$, e.g. occlusion between g_{t-1,k_1} and g_{t-1,k_2} is detected.
4. Randomly propagate the particles to enhance their diversities within the species according to the following transition model
$$x^{i,0}_{t,k} \sim \mathcal{N}(p^i_{t-1,k}, \Sigma_k)$$
where Σ_k is the covariance matrix of the Gaussian transition distribution for the *k*th object.
5. **for** $n = 1, 2, \ldots, T$ **do**
6. Carry out the PSO evolution for object k_1.
$$v^{i,n+1}_{t,k_1} = |r_1|(p^i_{t,k_1} - x^{i,n}_{t,k_1}) + |r_2|(g_{t,k_1} - x^{i,n}_{t,k_1}) + |r_3| F_{\overrightarrow{k_2,k_1}} + \epsilon$$
$$x^{i,n+1}_{t,k_1} = x^{i,n}_{t,k_1} + v^{i,n+1}_{t,k_1}$$
7. Evaluate the fitness values by the observation model and the interactive model.
$$f(x^{i,n+1}_{t,k_1}) = p(\tilde{o}_{t,k_1}|x^{i,n+1}_{t,k_1}) p(\hat{o}_{t,k_1}|x^{i,n+1}_{t,k_1}, g_{t,k_2})$$
8. Update the two best particles and the annealing parameter.
9. Carry out the similar procedure for object k_2 (other trackers are independently carried out without interactive part).
10. Check the convergence criteria: $f(g_{t,k_i})$ >*Th* and $f(g_{t,k_i}), i = 1, 2$ changes little from previous iteration.
11. If the convergence criteria for the object k_i is satisfied, **stop its iteration**;
12. **end for**
13. Update the appearance model based the visible parts and the corresponding reconstruction error.
14. Output the object states at time t: $\{g_{t,k}\}_{k=1}^M$
15. **end for**

[3] Here, we take pairwise occlusion as an example. The occlusion between three or more objects can be formulated similarly, and we only show how the tracking process is conducted on objects under occlusion. The trackers for unoccluded objects are conducted independently without the interaction part.

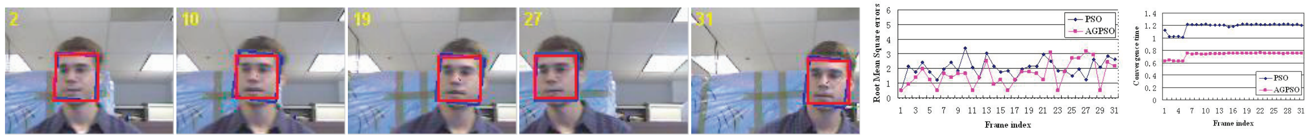

Figure 4. Column 1-5: tracking performances, column 6: root mean square error, column 7: convergence time (red: PSO, blue: AGPSO).

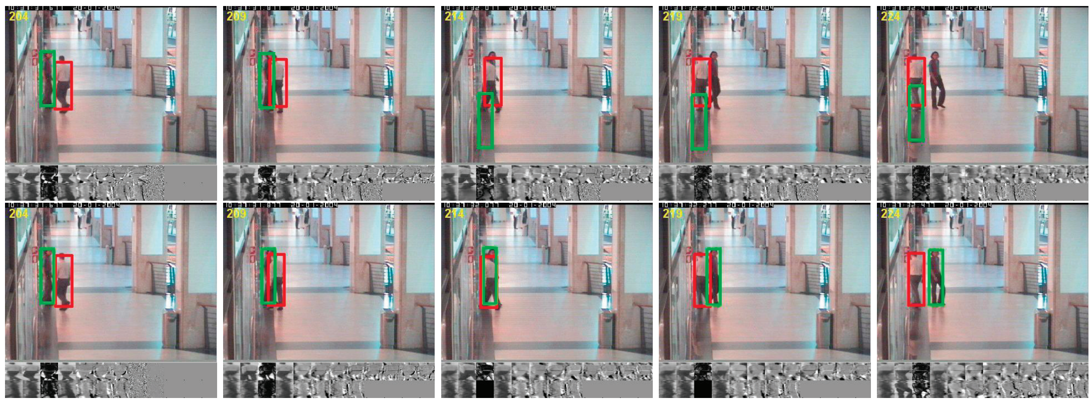

Figure 5. Tracking two walking men with occlusion (first row: stop updating for the appearance model during occlusion, second row: selective updating for the appearance model during occlusion) for frame #204, 209, 214, 219, 224

5. Experimental Results

In our implementation, each candidate image corresponding to a particle is rectified to a 20×20 patch, and the feature is a 400-dimension vector of gray level values subjected to zero-mean-unit-variance normalization. The above algorithm is implemented using Matlab on a P4-3.2G computer with 512M RAM.

5.1. PSO vs AGPSO

First, we conduct a comparison experiment between the traditional PSO [24] and AGPSO. Here, the particle number and the covariance matrix of the transition distribution are set to $\{N = 200, \Sigma = diag(8^2, 8^2, 0.02^2, 0.02^2, 0.002^2, 0.002^2)\}$. The same observation model is used for PSO and AGPSO. Fig. 4 shows the tracking performances of PSO and AGPSO on a fast moving face, and graphs of the RMSE (root mean square error) and convergence time, from which we can see that AGPSO can achieve a better tracking accuracy and a much faster convergence rate than traditional PSO. This is because the velocity part employed in Eq.(1) carries little information, while the annealing part in APSO iterations enhances the diversity of the particle set and its adaptive effect enables a fast convergence rate.

5.2. Multiple Object Tracking

In this section, we demonstrate three examples of tracking multiple objects with our method, and then give a summary of the experimental results.

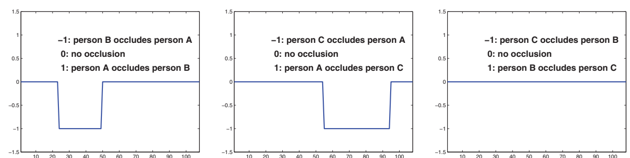

Figure 7. Recovered occlusion relationship in Example 2

5.2.1 Example 1

The video in this example contains two walking people with severe occlusion. We conduct a comparison experiment between two appearance updating strategies during occlusion: no updating and selective updating. As shown in Fig.5, we can see that no updating for the appearance model strategy fails to track the person being occluded at frame 211 and can not recover the track after occlusion. The reason is that no updating strategy does not capture the gradual appearance changes of the man being occluded, and thus the correspondence of pixels between the man and the subspace is not accurate, leading to the tracking failure. In contrary, our selective updating strategy can follow the two people throughout the occlusion and maintain the correct identities. This is because the appearance changes are gradually updated before the object is completely occluded, and all the existing visual evidences are utilized to successfully reacquire the man after the occlusion.

5.2.2 Example 2

In this example, we test our method with a video sequence from the PETS 2004 database which is an open database for the research on visual surveillance, available on http://homepages.inf.ed.ac.uk/rbf/CAVIAR/. To validate the claimed contributions of our method, we conduct a

(a) our algorithm

(b) Qu's work [9]

(c) Yang's work [12]

Figure 6. Tracking performances with occlusion for frame #422,455,465,480,488,501,518

Approaches		Yang's work	Qu's work	Our algorithm
Number of frames in which tracking is successful	Person A (red window)	80/101	80/108	108/108
	Person B (blue window)	108/108	108/108	108/108
	Person C (green window)	108/108	108/108	108/108
RMSE of Position (by pixels)	Person A (red window)	12.9768	11.5537	3.6145
	Person B (blue window)	5.4128	4.8482	3.3087
	Person C (green window)	15.2104	2.6483	2.6262

Table 1. Quantitative results of our approach and its comparison with Yang's and Qu's work

quantitative evaluation comparison with [9, 12] which share some similarities with our work, and furthermore, their work are respectively conducted in two influential frameworks: particle filter [25] and mean shift [26]. To give a convincing comparison, [9] is performed with the same appearance model and updating scheme. Fig. 6 illustrates key frames where three people are tracked through occlusion (Person A is tracked with a red window, person B is tracked with a green window, person C is tracked with a blue window), from which we can see that our algorithm handles the interaction and occlusion between different objects very well, while [9, 12] fail to track the object being occluded by an object with similar appearance, because modeling the species competition on the overlapping part and dealing it as a whole is more reasonable and robust. Meanwhile, our AGPSO framework is more probable to find the global optima than particle filter and mean shift. Fig. 7 shows the recovered occlusion relationships between different persons, where the x axis is the frame index, and the y is the occlusion relationship. As illustrated in Fig. 7, our method can correctly deduce the occlusion relationship based on the interactive likelihood, and the results support the phenomenon that the object with higher fitness value on the overlapping part is more likely to be the one occluding the other objects.

To further illustrate the advantages of our method, we conduct a quantitative evaluation comparison with [9, 12] in the following aspects: number of frames in which tracking is successful, RMSE (root mean square error) between the estimated position and the groundtruth[4]. Table 1 shows the quantitative comparison. It is clear that the [9, 12] fail to track person A at frame 501, when he is severely occluded by person C wearing the similar clothes, while our method using the species competition and repulsion model can prevent the coalescence problem and succeed in tracking throughout the sequence. Additionally, our method achieves the most accurate localization than other two methods.

5.2.3 Example 3

This video sequence is also from the PETS 2004 database, and it is more challenging since it contains five walking people with continual occlusion and interactions. Fig. 8 illustrates some key frames where five persons are tracked through the occlusion. As shown in Fig. 8, they are tracked well in the following sequences even though the occlusion simultaneously happens among the three persons at frames 277-340, from which we can see that our species competition and repulsion model are also effective for dealing with

[4]The object is not closely warped by the bounding box in the ground true which contains many background pixels. Initialized by such bounding boxes is not suitable for a subspace based tracking algorithm. Therefore, only the center position of each object in the groundtruth is used for evaluation.

Figure 8. Tracking people in a shopping center for frame #218,248,272,298,359,406

the occlusion among more than two objects. Besides the species competition and repulsion model, the selective updating of appearances during occlusion also provides a great contribution to maintain the correct tracking identities in this video sequence.

5.3. Summary

The underlying reasons for the above experimental results are discussed in this part. First, the species competition and repulsion force mechanism employed in our method provides a reasonable and effective solution to the interaction and occlusion problems in multiple object tracking. Second, the AGPSO framework is effective at searching for the optima, especially in a high dimension. Third, the carefully designed updating strategy can effectively accommodate the appearance changes while preventing the model from drifting away.

6. Conclusion

This paper makes an analogy between the tracking problem and the behavior of a flock of birds searching for food, and has proposed a species based sequential PSO (particle swarm optimization) algorithm for multi-object tracking, in which different species search for object (food) and track them once found. The occlusion between different objects is modeled as species competition and repulsion. In addition, we use an annealed Gaussian PSO algorithm which is more effective than previous PSO algorithms. Unlike the joint tracker, our approach decentralizes the joint tracker, and the individual trackers are conducted, each of which tries to maximize its visual evidence. Experimental results demonstrate the efficiency and effectiveness of our method.

Acknowledgment

We would like to thank Dr. Ming Yang for the help on implementing his algorithm. This work is partly supported by NSFC (Grant No. 60825204, 60672040) and the National 863 High-Tech R&D Program of China (Grant No. 2006AA01Z453, 2009AA01Z318).

References

[1] D. Reid, "An algorithm for tracking multiple targets", *IEEE Trans. on Automatic Control*, 24(6):843-854, 1979.

[2] Y. Bar-Shalom, T. Fortmann, and M. Scheffe, "Joint probabilistic data association for multiple targets in clutter", *In: Proc. Conf. on Information Sciences and Systems*, 1980.

[3] J. MacCormick and A. Blake, "A Probabilistic Exclusion Principle for Tracking Multiple Objects", *IJCV*, 39(1):57-71, 2000. 1

[4] M. Isard and J. MacCormick, "Bramble: A bayesian multipleblob tracker", *In: Proc. of ICCV*, pages 34-41, 2001. 1

[5] C. Rasmussen and G. Hager, "Probabilistic data association methods for tracking complex visual objects", *IEEE Trans. on PAMI*, 23(6):560-576, 2001.

[6] Z. Khan, T. Balch, and F. Dellaert, "An MCMC-based particle filter for tracking multiple interacting targets", *In: Proc. of ECCV*, pages 279-290, 2004. 1

[7] T. Yu and Y. Wu, "Collaborative tracking of multiple targets", *In: Proc. of CVPR*, pages 834-841, 2004. 1

[8] T. Zhao, R. Nevatia, "Tracking Multiple Humans in Crowded Environment", *In: Proc. of CVPR*, 2004

[9] W. Qu, D. Schonfeld, and M. Mohamed, "Real-Time Interactively Distributed Multi-Object Tracking Using a Magnetic- Inertia Potential Model", *In: Proc. of ICCV, pages 535-540*, 2005. 1, 4, 7

[10] H. Nguyen, Q. Ji, and A. Smeulders, "Robust multi-target tracking using spatio-temporal context", *In: Proc. of CVPR*, pages 578-585, 2006. 1

[11] P. Nillius, J. Sullivan, and S. Carlsson, "Multi-Target Tracking-CLinking Identities using Bayesian Network Inference", *In: Proc. of CVPR*, pages 2187-2194, 2006. 1

[12] M. Yang, T. Yu, and Y. Wu, "Game-Theoretic Multiple Target Tracking", *In: Proc. of ICCV*, 2007. 1, 4, 7

[13] Y. Jin and F. Mokhtarian, "Variational Particle Filter for Multi-Object Tracking", In: Proc. of ICCV, 2007. 1

[14] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera People Tracking with a Probabilistic Occupancy Map", *IEEE Trans. on PAMI*, 30(2):267-282, 2008.

[15] L. Zhang, Y. Li and R. Nevatia, "Global Data Association for Multi-Object Tracking Using Network Flows", *In: Proc. of CVPR*, 2008.

[16] A. Ess, B. Leibe, K. Schindler and L. V. Gool, "A Mobile Vision System for Robust Multi-Person Tracking", *In: Proc. of CVPR*, 2008.

[17] K. Ishiguro, T. Yamada and N. Ueda, "Simultaneous Clustering and Tracking Unknown Number of Objects", *In: Proc. of CVPR*, 2008.

[18] J. Kennedy and R. Eberhart, "Particle swarm optimization", *In: Proc. of IEEE International Conference on Neural Networks*, pages 1942-1948, 1995. 1, 2

[19] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in amultidimensional complex space", *IEEE Trans. on Evolutionary Computation*, 6(1):58-73, 2002. 1

[20] J. Lim, D. Ross, R. S. Lin, and M. H. Yang, "Incremental learning for visual tracking", *In Advances in Neural Information Processing Systems*, pp.793-800, 2004, The MIT Press. 3

[21] G. H. Golub and C. F. Van Loan, "Matrix Computations", *The Johns Hopkins University Press*, 1996. 3

[22] X. Zhang, W. Hu, S. Maybank, "A Smarter Particle Filter", *In: Proc. of ACCV*, 2009. 5

[23] L. Ingber, "Simulated Annealing: Practice Versus Theory", *Journal of Mathematical and Computer Modeling*, 18(11): 29-57, 1993. 5

[24] X. Zhang, W. Hu, S. Maybank, X. Li, and M. Zhu, "Sequential Particle Swarm Optimization for Visual Tracking", *In: Proc. of CVPR*, pages 1-8, 2008. 6

[25] M. Isard, and A. Blake, "Condensation: conditional density propagation for visual tracking", *International Journal of Computer Vision*, 29(1):5-28, 1998. 7

[26] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5): 234-240, 2003. 7