

An Introduction to Interacting Simulated Annealing

Juergen Gall, Bodo Rosenhahn, and Hans-Peter Seidel

Max-Planck Institute for Computer Science
Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany

Summary. Human motion capturing can be regarded as an optimization problem where one searches for the pose that minimizes a previously defined error function based on some image features. Most approaches for solving this problem use iterative methods like gradient descent approaches. They work quite well as long as they do not get distracted by local optima. We introduce a novel approach for global optimization that is suitable for the tasks as they occur during human motion capturing. We call the method interacting simulated annealing since it is based on an interacting particle system that converges to the global optimum similar to simulated annealing. We provide a detailed mathematical discussion that includes convergence results and annealing properties. Moreover, we give two examples that demonstrate possible applications of the algorithm, namely a global optimization problem and a multi-view human motion capturing task including segmentation, prediction, and prior knowledge. A quantitative error analysis also indicates the performance and the robustness of the interacting simulated annealing algorithm.

13.1 Introduction

13.1.1 Motivation

Optimization problems arise in many applications of computer vision. In pose estimation, e.g. [28], and human motion capturing, e.g. [31], functions are minimized at various processing steps. For example, the marker-less motion capture system [26] minimizes in a first step an energy function for the segmentation. In a second step, correspondences between the segmented image and a 3D model are established. The optimal pose is then estimated by minimizing the error given by the correspondences. These optimization problems also occur, for instance, in model fitting [17, 31]. The problems are mostly solved by iterative methods as gradient descent approaches. The methods work very well as long as the starting point is near the global optimum, however, they get easily stuck in a local optimum. In order to deal with it, several random selected starting points are used and the best solution is selected in the hope that at least one of them is near enough to the global optimum, cf. [26]. Although it improves the results in many cases, it does not ensure that the global optimum is found.

In this chapter, we introduce a global optimization method based on an interacting particle system that overcomes the dilemma of local optima and that is suitable for the optimization problems as they arise in human motion capturing. In contrast to many other optimization algorithms, a distribution instead of a single value is approximated by a particle representation similar to particle filters [10]. This property is beneficial, particularly, for tracking where the right parameters are not always exact at the global optimum depending on the image features that are used.

13.1.2 Related Work

A popular global optimization method inspired by statistical mechanics is known as simulated annealing [14, 18]. Similar to our approach, a function $V \geq 0$ interpreted as energy is minimized by means of an unnormalized *Boltzmann-Gibbs measure* that is defined in terms of V and an inverse temperature $\beta > 0$ by

$$g(dx) = \exp(-\beta V(x)) \lambda(dx), \quad (13.1)$$

where λ is the Lebesgue measure. This measure has the property that the probability mass concentrates at the global minimum of V as $\beta \rightarrow \infty$.

The key idea behind simulated annealing is taking a random walk through the search space while β is successively increased. The probability of accepting a new value in the space is given by the Boltzmann-Gibbs distribution. While values with less energy than the current value are accepted with probability one, the probability that values with higher energy are accepted decreases as β increases. Other related approaches are fast simulated annealing [30] using a Cauchy-Lorentz distribution and generalized simulated annealing [32] based on Tsallis statistics.

Interacting particle systems [19] approximate a distribution of interest by a finite number of weighted random variables $X^{(i)}$ called particles. Provided that the weights $\Pi^{(i)}$ are normalized such that $\sum \Pi^{(i)} = 1$, the set of weighted particles determines a random probability measures by

$$\sum_{i=1}^n \Pi^{(i)} \delta_{X^{(i)}}. \quad (13.2)$$

Depending on the weighting function and the distribution of the particles, the measure converges to a distribution η as n tends to infinity. When the particles are identically independently distributed according to η and uniformly weighted, i.e. $\Pi^{(i)} = 1/n$, the convergence follows directly from the law of large numbers [3].

Interacting particle systems are mostly known in computer vision as particle filter [10] where they are applied for solving non-linear, non-Gaussian filtering problems. However, these systems also apply for trapping analysis, evolutionary algorithms, statistics [19], and optimization as we demonstrate in this chapter. They usually consist of two steps as illustrated in Figure 13.1. During a selection step, the particles are weighted according to a weighting function and then resampled with respect to their weights, where particles with a great weight generate more offspring than particles with lower weight. In a second step, the particles mutate or are diffused.

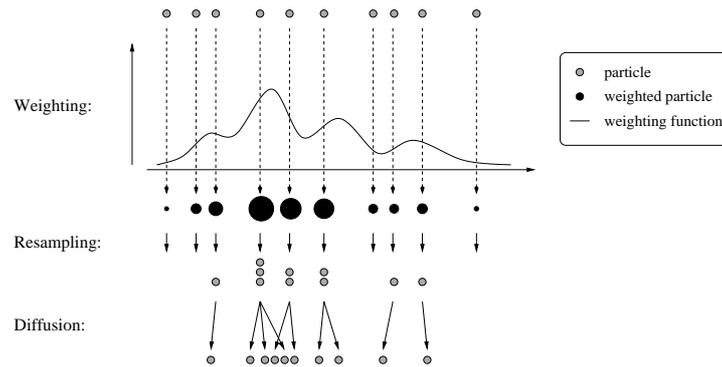


Fig. 13.1. Operation of an interacting particle system. After weighting the particles (*black circles*), the particles are resampled and diffused (*gray circles*).

13.1.3 Interaction and Annealing

Simulated annealing approaches are designed for global optimization, i.e. for searching the global optimum in the entire search space. Since they are not capable of focusing the search on some regions of interest in dependency on the previous visited values, they are not suitable for tasks in human motion capturing. Our approach, in contrast, is based on an interacting particle system that uses Boltzmann-Gibbs measures (13.1) similar to simulated annealing. This combination ensures not only the annealing property as we will show, but also exploits the distribution of the particles in the space as measure for the uncertainty in an estimate. The latter allows an automatic adaption of the search on regions of interest during the optimization process. The principle of the annealing effect is illustrated in Figure 13.2.

A first attempt to fuse interaction and annealing strategies for human motion capturing has become known as annealed particle filter [9]. Even though the heuristic is not based on a mathematical background, it already indicates the potential of such combination. Indeed, the annealed particle filter can be regarded as a special case of interacting simulated annealing where the particles are predicted for each frame by a stochastic process, see Section 13.3.1.

13.1.4 Outline

The interacting annealing algorithm is introduced in Section 13.3.1 and its asymptotic behavior is discussed in Section 13.3.2. The given convergence results are based on Feynman-Kac models [19] which are outlined in Section 13.2. Since a general treatment including proofs is out of the scope of this introduction, we refer the interested reader to [11] or [19]. While our approach is evaluated for a standard global optimization problem in Section 13.4.1, Section 13.4.2 demonstrates the performance of interacting simulated annealing in a complete marker-less human motion capture system that includes segmentation, pose prediction and prior knowledge.

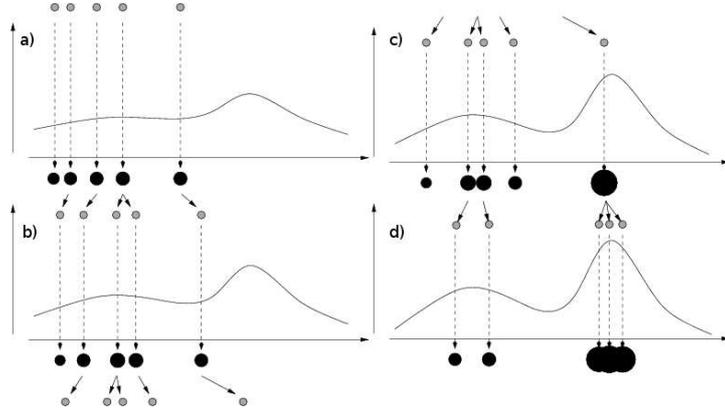


Fig. 13.2. Illustration of the annealing effect with three runs. Due to annealing, the particles migrate towards the global maximum without getting stuck in the local maximum.

13.1.5 Notations

We always regard E as a subspace of R^d , and let $\mathcal{B}(E)$ denote its Borel σ -algebra. $B(E)$ denotes the set of bounded measurable functions, δ_x is the Dirac measure concentrated in $x \in E$, $\| \cdot \|_2$ is the Euclidean norm, and $\| \cdot \|_\infty$ is the well-known supremum norm. Let $f \in B(E)$, μ be a measure on E , and let K be a Markov kernel on E^1 . We write

$$\langle \mu, f \rangle = \int_E f(x) \mu(dx), \quad \langle \mu, K \rangle(B) = \int_E K(x, B) \mu(dx) \quad \text{for } B \in \mathcal{B}(E).$$

Furthermore, $U[0, 1]$ denotes the uniform distribution on the interval $[0, 1]$ and

$$\text{osc}(\varphi) := \sup_{x, y \in E} \{|\varphi(x) - \varphi(y)|\}. \tag{13.3}$$

is an upper bound for the oscillations of f .

13.2 Feynman-Kac Model

Let $(X_t)_{t \in \mathbb{N}_0}$ be an E -valued Markov process with family of transition kernels $(K_t)_{t \in \mathbb{N}_0}$ and initial distribution η_0 . We denote by P_{η_0} the distribution of the Markov process, i.e. for $t \in \mathbb{N}_0$,

$$P_{\eta_0}(d(x_0, x_1, \dots, x_t)) = K_{t-1}(x_{t-1}, dx_t) \dots K_0(x_0, dx_1) \eta_0(dx_0),$$

¹ A Markov kernel is a function $K : E \times \mathcal{B}(E) \rightarrow [0, \infty]$ such that $K(\cdot, B)$ is $\mathcal{B}(E)$ -measurable $\forall B$ and $K(x, \cdot)$ is a probability measure $\forall x$. An example of a Markov kernel is given in Equation (13.12). For more details on probability theory and Markov kernels, we refer to [3].

and by $E_{\eta_0} [\cdot]$ the expectation with respect to P_{η_0} . The sequence of distributions $(\eta_t)_{t \in \mathbb{N}_0}$ on E defined for any $\varphi \in B(E)$ and $t \in \mathbb{N}_0$ as

$$\langle \eta_t, \varphi \rangle := \frac{\langle \gamma_t, \varphi \rangle}{\langle \gamma_t, 1 \rangle}, \quad \langle \gamma_t, \varphi \rangle := E_{\eta_0} \left[\varphi(X_t) \exp \left(- \sum_{s=0}^{t-1} \beta_s V(X_s) \right) \right],$$

is called the *Feynman-Kac model* associated with the pair $(\exp(-\beta_t V), K_t)$. The Feynman-Kac model as defined above satisfies the recursion relation

$$\eta_{t+1} = \langle \Psi_t(\eta_t), K_t \rangle, \tag{13.4}$$

where the *Boltzmann-Gibbs transformation* Ψ_t is defined by

$$\Psi_t(\eta_t)(dy_t) = \frac{E_{\eta_0} \left[\exp \left(- \sum_{s=0}^{t-1} \beta_s V(X_s) \right) \right]}{E_{\eta_0} \left[\exp \left(- \sum_{s=0}^t \beta_s V(X_s) \right) \right]} \exp(-\beta_t V_t(y_t)) \eta_t(dy_t).$$

The particle approximation of the flow (13.4) depends on a chosen family of Markov transition kernels $(K_{t, \eta_t})_{t \in \mathbb{N}_0}$ satisfying the compatibility condition

$$\langle \Psi_t(\eta_t), K_t \rangle := \langle \eta_t, K_{t, \eta_t} \rangle.$$

A family $(K_{t, \eta_t})_{t \in \mathbb{N}_0}$ of kernels is not uniquely determined by these conditions. As in [19, Chapter 2.5.3], we choose

$$K_{t, \eta_t} = S_{t, \eta_t} K_t, \tag{13.5}$$

where

$$S_{t, \eta_t}(x_t, dy_t) = \epsilon_t \exp(-\beta_t V_t(x_t)) \delta_{x_t}(dy_t) + (1 - \epsilon_t \exp(-\beta_t V_t(x_t))) \Psi_t(\eta_t)(dy_t), \tag{13.6}$$

with $\epsilon_t \geq 0$ and $\epsilon_t \|\exp(-\beta_t V)\|_\infty \leq 1$. The parameters ϵ_t may depend on the current distribution η_t .

13.3 Interacting Simulated Annealing

Similar to simulated annealing, one can define an annealing scheme $0 \leq \beta_0 \leq \beta_1 \leq \dots \leq \beta_t$ in order to search for the global minimum of an energy function V . Under some conditions that will be stated in Section 13.3.2, the flow of the Feynman-Kac distribution becomes concentrated in the region of global minima of V as t goes to infinity. Since it is not possible to sample from the distribution directly, the flow is approximated by a particle set as it is done by a particle filter. We call the algorithm for the flow approximation *interacting simulated annealing (ISA)*.

13.3.1 Algorithm

The particle approximation for the Feynman-Kac model is completely described by the Equation (13.5). The particle system is initialized by n identically, independently distributed random variables $X_0^{(i)}$ with common law η_0 determining the

random probability measure $\eta_0^n := \sum_{i=1}^n \delta_{X_0^{(i)}}/n$. Since K_{t,η_t} can be regarded as the composition of a pair of selection and mutation Markov kernels, we split the transitions into the following two steps

$$\eta_t^n \xrightarrow{\text{Selection}} \tilde{\eta}_t^n \xrightarrow{\text{Mutation}} \eta_{t+1}^n,$$

where

$$\eta_t^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_t^{(i)}}, \quad \tilde{\eta}_t^n := \frac{1}{n} \sum_{i=1}^n \delta_{\tilde{X}_t^{(i)}}.$$

During the selection step each particle $X_t^{(i)}$ evolves according to the Markov transition kernel $S_{t,\eta_t^n}(X_t^{(i)}, \cdot)$. That means $X_t^{(i)}$ is accepted with probability

$$\epsilon_t \exp(-\beta_t V(X_t^{(i)})), \tag{13.7}$$

and we set $\tilde{X}_t^{(i)} = X_t^{(i)}$. Otherwise, $\tilde{X}_t^{(i)}$ is randomly selected with distribution

$$\sum_{i=1}^n \frac{\exp(-\beta_t V(X_t^{(i)}))}{\sum_{j=1}^n \exp(-\beta_t V(X_t^{(j)}))} \delta_{X_t^{(i)}}.$$

The mutation step consists in letting each selected particle $\tilde{X}_t^{(i)}$ evolve according to the Markov transition kernel $K_t(\tilde{X}_t^{(i)}, \cdot)$.

Algorithm 6 Interacting Simulated Annealing Algorithm

Requires: parameters $(\epsilon_t)_{t \in \mathbb{N}_0}$, number of particles n , initial distribution η_0 , energy function V , annealing scheme $(\beta_t)_{t \in \mathbb{N}_0}$ and transitions $(K_t)_{t \in \mathbb{N}_0}$

1. Initialization
 - Sample $x_0^{(i)}$ from η_0 for all i
 2. Selection
 - Set $\pi^{(i)} \leftarrow \exp(-\beta_t V(x_t^{(i)}))$ for all i
 - For i from 1 to n :
 - Sample κ from $U[0, 1]$
 - If $\kappa \leq \epsilon_t \pi^{(i)}$ then
 - ★ Set $\tilde{x}_t^{(i)} \leftarrow x_t^{(i)}$
 - Else
 - ★ Set $\tilde{x}_t^{(i)} \leftarrow x_t^{(j)}$ with probability $\frac{\pi^{(j)}}{\sum_{k=1}^n \pi^{(k)}}$
 3. Mutation
 - Sample $x_{t+1}^{(i)}$ from $K_t(\tilde{x}_t^{(i)}, \cdot)$ for all i and go to step 2
-

There are several ways to choose the parameter ϵ_t of the *selection kernel* (13.6) that defines the resampling procedure of the algorithm, cf. [19]. If

$$\epsilon_t := 0 \quad \forall t, \tag{13.8}$$

the selection can be done by multinomial resampling. Provided that²

² The inequality satisfies the condition $\epsilon_t \|\exp(-\beta_t V)\|_\infty \leq 1$ for Equation (13.6).

$$n \geq \sup_t (\exp(\beta_t \text{osc}(V))),$$

another selection kernel is given by

$$\epsilon_t(\eta_t) := \frac{1}{n \langle \eta_t, \exp(-\beta_t V) \rangle}. \tag{13.9}$$

In this case the expression $\epsilon_t \pi^{(i)}$ in Algorithm 6 is replaced by $\pi^{(i)} / \sum_{k=1}^n \pi^{(k)}$. A third kernel is determined by

$$\epsilon_t(\eta_t) := \frac{1}{\inf \{y \in \mathbb{R} : \eta_t(\{x \in E : \exp(-\beta_t V(x)) > y\}) = 0\}}, \tag{13.10}$$

yielding the expression $\pi^{(i)} / \max_{1 \leq k \leq n} \pi^{(k)}$ instead of $\epsilon_t \pi^{(i)}$.

Pierre del Moral showed in [19, Chapter 9.4] that for any $t \in \mathbb{N}_0$ and $\varphi \in B(E)$ the sequence of random variables

$$\sqrt{n}(\langle \eta_t^n, \varphi \rangle - \langle \eta_t, \varphi \rangle)$$

converges in law to a Gaussian random variable W when the selection kernel (13.6) is used to approximate the flow (13.4). Moreover, it turns out that when (13.9) is chosen, the variance of W is strictly smaller than in the case with $\epsilon_t = 0$.

We remark that the annealed particle filter [9] relies on interacting simulated annealing with $\epsilon_t = 0$. The operation of the method is illustrated by

$$\eta_t^n \xrightarrow{\text{Prediction}} \hat{\eta}_{t+1}^n \xrightarrow{\text{ISA}} \eta_{t+1}^n.$$

The *ISA* is initialized by the predicted particles $\hat{X}_{t+1}^{(i)}$ and performs M times the selection and mutation steps. Afterwards the particles $X_{t+1}^{(i)}$ are obtained by an additional selection. This shows that the annealed particle filter uses a simulated annealing principle to locate the global minimum of a function V at each time step.

13.3.2 Convergence

This section discusses the asymptotic behavior of the interacting simulated annealing algorithm. For this purpose, we introduce some definitions in accordance with [19] and [15].

Definition 1. A kernel K on E is called *mixing* if there exists a constant $0 < \varepsilon < 1$ such that

$$K(x_1, \cdot) \geq \varepsilon K(x_2, \cdot) \quad \forall x_1, x_2 \in E. \tag{13.11}$$

The condition can typically only be established when $E \subset \mathbb{R}^d$ is a bounded subset, which is the case in many applications like human motion capturing. For example the (bounded) Gaussian distribution on E

$$K(x, B) := \frac{1}{Z} \int_B \exp\left(-\frac{1}{2}(x-y)^T \Sigma^{-1}(x-y)\right) dy, \tag{13.12}$$

where $Z := \int_E \exp(-\frac{1}{2}(x-y)^T \Sigma^{-1}(x-y)) dy$, is mixing if and only if E is bounded. Moreover, a Gaussian with a high variance satisfies the mixing condition with a larger ε than a Gaussian with lower variance.

Definition 2. *The Dobrushin contraction coefficient of a kernel K on E is defined by*

$$\beta(K) := \sup_{x_1, x_2 \in E} \sup_{B \in \mathcal{B}(E)} |K(x_1, B) - K(x_2, B)|. \tag{13.13}$$

Furthermore, $\beta(K) \in [0, 1]$ and $\beta(K_1 K_2) \leq \beta(K_1) \beta(K_2)$.

When the kernel M is a composition of several mixing Markov kernels, i.e. $M := K_s K_{s+1} \dots K_t$, and each kernel K_k satisfies the mixing condition for some ε_k , the Dobrushin contraction coefficient can be estimated by $\beta(M) \leq \prod_{k=s}^t (1 - \varepsilon_k)$. The asymptotic behavior of the interacting simulated annealing algorithm is affected by the convergence of the flow of the Feynman-Kac distribution (13.4) to the region of global minima of V as t tends to infinity and by the convergence of the particle approximation to the Feynman-Kac distribution at each time step t as the number of particles n tends to infinity.

Convergence of the flow

We suppose that $K_t = K$ is a Markov kernel satisfying the mixing condition (13.11) for an $\varepsilon \in (0, 1)$ and $\text{osc}(V) < \infty$. A time mesh is defined by

$$t(n) := n(1 + \lfloor c(\varepsilon) \rfloor) \quad c(\varepsilon) := (1 - \ln(\varepsilon/2))/\varepsilon^2 \quad \text{for } n \in \mathbb{N}_0. \tag{13.14}$$

Let $0 \leq \beta_0 \leq \beta_1 \dots$ be an annealing scheme such that $\beta_t = \beta_{t(n+1)}$ is constant in the interval $(t(n), t(n+1)]$. Furthermore, we denote by $\tilde{\eta}_t$ the Feynman-Kac distribution after the selection step, i.e. $\tilde{\eta}_t = \Psi_t(\eta_t)$. According to [19, Proposition 6.3.2], we have

Theorem 1. *Let $b \in (0, 1)$ and $\beta_{t(n+1)} = (n + 1)^b$. Then for each $\delta > 0$*

$$\lim_{n \rightarrow \infty} \tilde{\eta}_{t(n)}(V \geq V_* + \delta) = 0,$$

where $V_* = \sup\{v \geq 0; V \geq v \text{ a.e.}\}$.

The rate of convergence is $d/n^{(1-b)}$ where d is increasing with respect to b and $c(\varepsilon)$ but does not depend on n as given in [19, Theorem 6.3.1]. This theorem establishes that the flow of the Feynman-Kac distribution $\tilde{\eta}_t$ becomes concentrated in the region of global minima as $t \rightarrow +\infty$.

Convergence of the particle approximation

Del Moral established the following convergence theorem [19, Theorem 7.4.4].

Theorem 2. *For any $\varphi \in B(E)$,*

$$E_{\eta_0} [|\langle \eta_{t+1}^n, \varphi \rangle - \langle \eta_{t+1}, \varphi \rangle|] \leq \frac{2 \text{osc}(\varphi)}{\sqrt{n}} \left(1 + \sum_{s=0}^t r_s \beta(M_s) \right),$$

where

$$r_s := \exp \left(\text{osc}(V) \sum_{r=s}^t \beta_r \right),$$

$$M_s := K_s K_{s+1} \dots K_t,$$

for $0 \leq s \leq t$.

Assuming that the kernels K_s satisfy the mixing condition with ε_s , we get a rough estimate for the number of particles

$$n \geq \frac{4 \operatorname{osc}(\varphi)^2}{\delta^2} \left(1 + \sum_{s=0}^t \left\{ \exp \left(\operatorname{osc}(V) \sum_{r=s}^t \beta_r \right) \prod_{k=s}^t (1 - \varepsilon_k) \right\} \right)^2 \quad (13.15)$$

needed to achieve a mean error less than a given $\delta > 0$.

Optimal transition kernel

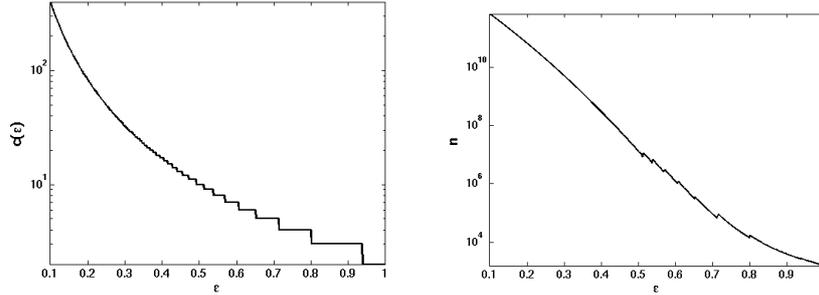


Fig. 13.3. Impact of the mixing condition satisfied for $\varepsilon_s = \varepsilon$. **Left:** Parameter $c(\varepsilon)$ of the time mesh (13.14). **Right:** Rough estimate for the number of particles needed to achieve a mean error less than $\delta = 0.1$.

The mixing condition is not only essential for the convergence result of the flow as stated in Theorem 1 but also influences the time mesh by the parameter ε . In view of Equation (13.14), kernels with ε close to 1 are preferable, e.g. Gaussian kernels on a bounded set with a very high variance. The right hand side of (13.15) can also be minimized if Markov kernels K_s are chosen such that the mixing condition is satisfied for a ε_s close to 1, as shown in Figure 13.3. However, we have to consider two facts. First, the inequality in Theorem 2 provides an upper bound of the accumulated error of the particle approximation up to time $t + 1$. It is clear that the accumulation of the error is reduced when the particles are highly diffused, but it also means that the information carried by the particles from the previous time steps is mostly lost by the mutation. Secondly, we cannot sample from the measure $\tilde{\eta}_t$ directly, instead we approximate it by n particles. Now the following problem arises. The mass of the measure concentrates on a small region of E on one hand and, on the other hand, the particles are spread over E if ε is large. As a result we get a degenerated system where the weights of most of the particles are zero and thus the global minima are estimated inaccurately, particularly for small n . If we choose a kernel with small ε in contrast, the convergence rate of the flow is very slow. Since neither of them is suitable in practice, we suggest a *dynamic variance scheme* instead of a fixed kernel K .

It can be implemented by Gaussian kernels K_t with covariance matrices Σ_t proportional to the sample covariance after resampling. That is, for a constant $c > 0$,

$$\Sigma_t := \frac{c}{n-1} \sum_{i=1}^n (x_t^{(i)} - \mu_t)_\rho (x_t^{(i)} - \mu_t)_\rho^T, \quad \mu_t := \frac{1}{n} \sum_{i=1}^n x_t^{(i)}, \quad (13.16)$$

where $((x)_\rho)_k = \max(x_k, \rho)$ for a $\rho > 0$. The value ρ ensures that the variance does not become zero. The elements off the diagonal are usually set to zero, in order to reduce computation time.

Optimal parameters

The computation cost of the interacting simulated annealing algorithm with n particles and T annealing runs is $O(nT)$, where

$$n_T := n \cdot T. \quad (13.17)$$

While more particles give a better particle approximation of the Feynman-Kac distribution, the flow becomes more concentrated in the region of global minima as the number of annealing runs increases. Therefore, finding the optimal values is a trade-off between the convergence of the flow and the convergence of the particle approximation provided that n_T is fixed.

Another important parameter of the algorithm is the annealing scheme. The scheme given in Theorem 1 ensures convergence for any energy function V — even for the worst one in the sense of optimization — as long as $\text{osc}(V) < \infty$ but is too slow for most applications, as it is the case for simulated annealing. In our experiments the schemes

$$\beta_t = \ln(t+b) \quad \text{for some } b > 1 \quad (\text{logarithmic}), \quad (13.18)$$

$$\beta_t = (t+1)^b \quad \text{for some } b \in (0,1) \quad (\text{polynomial}) \quad (13.19)$$

performed well. Note that in contrast to the time mesh (13.14) the schemes are not anymore constant on a time interval.

Even though a complete evaluation of the various parameters is out of the scope of this introduction, the examples given in the following section demonstrate settings that perform well, in particular for human motion capturing.

13.4 Examples

13.4.1 Global Optimization

The Ackley function [2, 1]

$$f(x) = -20 \exp \left(-0.2 \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2} \right) - \exp \left(\frac{1}{d} \sum_{i=1}^d \cos(2\pi x_i) \right) + 20 + e$$

is a widely used multimodal test function for global optimization algorithms. As one can see from Figure 13.4, the function has a global minimum at $(0,0)$ that is surrounded by several local minima. The problem consists of finding the global

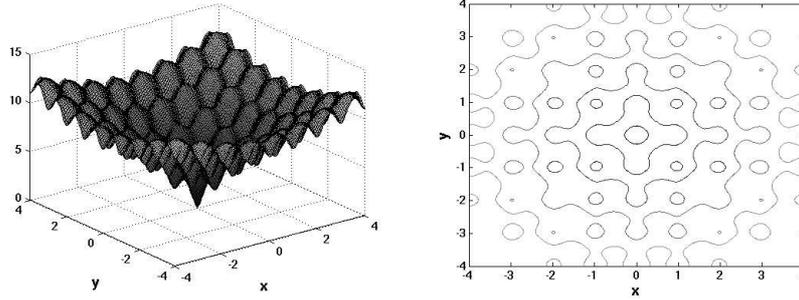


Fig. 13.4. Ackley function. Unique global minimum at $(0,0)$ with several local minima around it.

minimum in a bounded subspace $E \subset \mathbb{R}^d$ with an error less than a given $\delta > 0$ where the initial distribution is the uniform distribution on E .

In our experiments, the maximal number of time steps were limited by 999, and we set $E = [-4, 4] \times [-4, 4]$ and $\delta = 10^{-3}$. The interacting simulated annealing algorithm was stopped when the Euclidean distance between the global minimum and its estimate was less than δ or when the limit of time steps was exceeded. All simulations were repeated 50 times and the average number of time steps needed by *ISA* was used for evaluating the performance of the algorithm. Depending on the chosen selection kernel (13.8), (13.9), and (13.10), we write ISA_{S1} , ISA_{S2} , and ISA_{S3} , respectively.

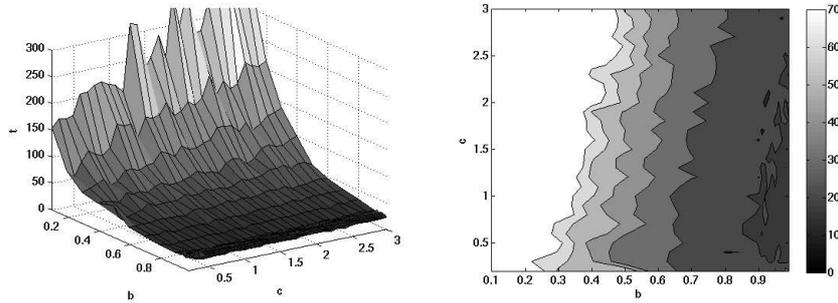


Fig. 13.5. Average time steps needed to find global minimum with error less than 10^{-3} with respect to the parameters b and c .

Using a polynomial annealing scheme (13.19), we evaluated the average time steps needed by the ISA_{S1} with 50 particles to find the global minimum of the Ackley function. The results with respect to the parameter of the annealing scheme, $b \in [0.1, 0.999]$, and the parameter of the dynamic variance scheme, $c \in [0.1, 3]$, are given in Figure 13.5. The algorithm performed best with a fast increasing annealing scheme, i.e. $b > 0.9$, and with c in the range $0.5 - 1.0$. The plots in Figure 13.5

also reveal that the annealing scheme has greater impact on the performance than the factor c . When the annealing scheme increases slowly, i.e. $b < 0.2$, the global minimum was actually not located within the given limit for all 50 simulations.

	Ackley			Ackley with noise		
	ISA_{S1}	ISA_{S2}	ISA_{S3}	ISA_{S1}	ISA_{S2}	ISA_{S3}
b	0.993	0.987	0.984	0.25	0.35	0.27
c	0.8	0.7	0.7	0.7	0.7	0.9
t	14.34	15.14	14.58	7.36	7.54	7.5

Table 13.1. Parameters b and c with lowest average time t for different selection kernels.

The best results with parameters b and c for ISA_{S1} , ISA_{S2} , and ISA_{S3} are listed in Table 13.1. The optimal parameters for the three selection kernels are quite similar and the differences of the average time steps are marginal.

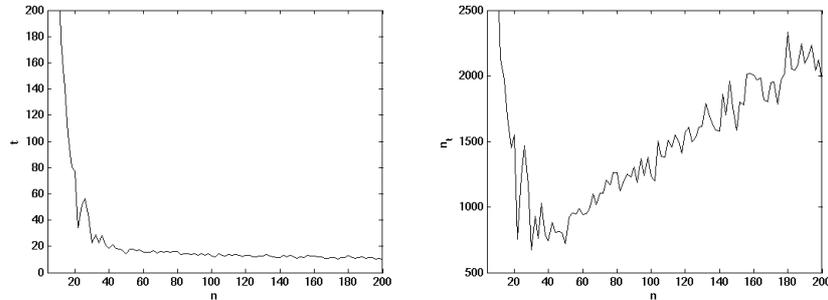


Fig. 13.6. **Left:** Average time steps needed to find global minimum with respect to number of particles. **Right:** Computation cost.

In a second experiment, we fixed the parameters b and c , where we used the values from Table 13.1, and varied the number of particles in the range 4 – 200 with step size 2. The results for ISA_{S1} are shown in Figure 13.6. While the average of time steps declines rapidly for $n \leq 20$, it is hardly reduced for $n \geq 40$. Hence, n_t and thus the computation cost are lowest in the range 20 – 40. This shows that a minimum number of particles are required to achieve a success rate of 100%, i.e., the limit was not exceeded for all simulations. In this example, the success rate was 100% for $n \geq 10$. Furthermore, it indicates that the average of time steps is significantly higher for n less than the optimal number of particles. The results for ISA_{S1} , ISA_{S2} , and ISA_{S3} are quite similar. The best results are listed in Table 13.2.

The ability of dealing with noisy energy functions is one of the strength of ISA as we will demonstrate. This property is very usefull for applications where the measurement of the energy of a particle is distorted by noise. On the left hand side

	Ackley			Ackley with noise		
	ISA_{S1}	ISA_{S2}	ISA_{S3}	ISA_{S1}	ISA_{S2}	ISA_{S3}
n	30	30	28	50	50	26
t	22.4	20.3	21.54	7.36	7.54	12.54
n_t	672	609	603.12	368	377	326.04

Table 13.2. Number of particles with lowest average computation cost for different selection kernels.

of Figure 13.7, the Ackley function is distorted by Gaussian noise with standard deviation 0.5, i.e.,

$$f_W(x) := \max \{0, f(x) + W\}, \quad W \sim N(0, 0.5^2).$$

As one can see, the noise deforms the shape of the function and changes the region of global minima. In our experiments, the ISA was stopped when the true global minimum at $(0, 0)$ was found with an accuracy of $\delta = 0.01$.

For evaluating the parameters b and c , we set $n = 50$. As shown on the right hand side of Figure 13.7, the best results were obtained by annealing schemes with $b \in [0.22, 0.26]$ and $c \in [0.6, 0.9]$. In contrast to the undistorted Ackley function, annealing schemes that increase slowly performed better than the fast one. Indeed, the success rate dropped below 100% for $b \geq 0.5$. The reason is obvious from the left hand side of Figure 13.7. Due to the noise, the particles are more easily distracted and a fast annealing scheme diminishes the possibility of escaping from the local minima. The optimal parameters for the dynamic variance scheme are hardly affected by the noise.

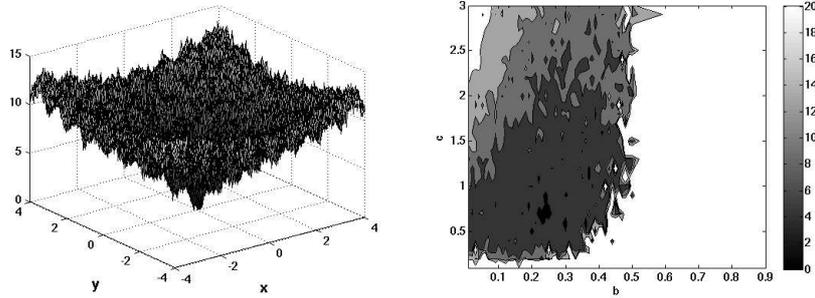


Fig. 13.7. **Left:** Ackley function distorted by Gaussian noise with standard deviation 0.5. **Right:** Average time steps needed to find global minimum with error less than 10^{-2} with respect to the parameters b and c .

The best parameters for ISA_{S1} , ISA_{S2} , and ISA_{S3} are listed in the Tables 13.1 and 13.2. Except for ISA_{S3} , the optimal number of particles is higher than it is the case for the simulations without noise. The minimal number of particles to achieve

a success rate of 100% also increased, e.g. 28 for ISA_{S1} . We remark that ISA_{S3} required the least number of particles for a complete success rate, namely 4 for the undistorted energy function and 22 in the noisy case.

We finish this section by illustrating two examples of energy function where the dynamic variance schemes might not be suitable. On the left hand side of Figure 13.8, an energy function with shape similar to the Ackley function is drawn. The dynamic variance schemes perform well for this type of function with a unique global minimum with several local minima around it. Due to the scheme, the search focuses on the region near the global minimum after some time steps. The second function, see Figure 8(b), has several, widely separated global minima yielding a high variance of the particles even in the case that the particles are near to the global minima. Moreover, when the region of global minima is regarded as a sum of Dirac measures, the mean is not essentially a global minimum. In the last example shown on the right hand side of Figure 13.8, the global minimum is a small peak far away from a broad basin with a local minimum. When all particles fall into the basin, the dynamic variance schemes focus the search on the region near the local minimum and it takes a long time to discover the global minimum.

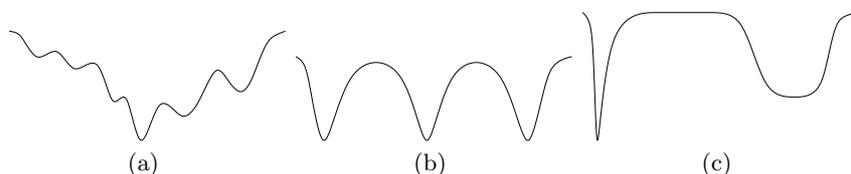


Fig. 13.8. Different cases of energy functions. **(a)** Optimal for dynamic variance schemes. An unique global minimum with several local minima around it. **(b)** Several global minima that are widely separated. This yields a high variance even in the case that the particles are near to the global minima. **(c)** The global minimum is a small peak far away from a broad basin. When all particles fall into the basin, the dynamic variance schemes focus the search on the basin.

In most optimization problems arising in the field of computer vision, however, the first case occurs where the dynamic variance schemes perform well. One application is human motion capturing which we will discuss in the next section.

13.4.2 Human Motion Capture

In our second experiment, we apply the interacting simulated annealing algorithm to model-based 3D tracking of the lower part of a human body, see Figure 13.9(a). This means that the 3D *rigid body motion (RBM)* and the joint angles, also called the *pose*, are estimated by exploiting the known 3D model of the tracked object. The mesh model illustrated in Figure 13.9(d) has 18 *degrees of freedom (DoF)*, namely 6 for the rigid body motion and 12 for the joint angles of the hip, knees, and feet. Although a marker-less motion capture system is discussed, markers are also stucked to the target object in order to provide a quantitative comparison with a commercial marker based system.



Fig. 13.9. From left to right: (a) Original image. (b) Silhouette. (c) Estimated pose. (d) 3D model.

Using the extracted silhouette as shown in Figure 13.9(b), one can define an energy function V which describes the difference between the silhouette and an estimated pose. The pose that fits the silhouette best takes the global minimum of the energy function, which is searched by the *ISA*. The estimated pose projected onto the image plane is displayed in Figure 13.9(c).

Pose representation

There are several ways to represent the pose of an object, e.g. Euler angles, quaternions [16], twists [20], or the axis-angle representation. The *ISA* requires from the representation that primarily the mean but also the variance can be at least well approximated. For this purpose, we have chosen the axis-angle representation of the absolute rigid body motion M given by the 6D vector $(\theta\omega, t)$ with

$$\omega = (\omega_1, \omega_2, \omega_3), \quad \|\omega\|_2 = 1 \quad \text{and} \quad t = (t_1, t_2, t_3).$$

Using the exponential, M is expressed by

$$M = \begin{pmatrix} \exp(\theta\hat{\omega}) & t \\ 0 & 1 \end{pmatrix}, \quad \hat{\omega} = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}. \quad (13.20)$$

While t is the absolute position in the world coordinate system, the rotation vector $\theta\omega$ describes a rotation by an angle $\theta \in \mathbb{R}$ about the rotation axis ω . The function $\exp(\theta\hat{\omega})$ can be efficiently computed by the Rodriguez formula [20].

Given a rigid body motion defined by a rotation matrix $R \in SO(3)$ and a translation vector $t \in \mathbb{R}^3$, the rotation vector is constructed according to [20] as follows: When R is the identity matrix, θ is set to 0. For the other case, θ and the rotation axis ω are given by

$$\theta = \cos^{-1} \left(\frac{\text{trace}(R) - 1}{2} \right), \quad \omega = \frac{1}{2 \sin(\theta)} \begin{pmatrix} r_{32} - r_{23} \\ r_{13} - r_{31} \\ r_{21} - r_{12} \end{pmatrix}. \quad (13.21)$$

We write $\log(R)$ for the inverse mapping of the exponential.

The mean of a set of rotations r_i in the axis-angle representation can be computed by using the exponential and the logarithm as described in [22, 23]. The idea is to find a geodesic on the Riemannian manifold determined by the set of 3D rotations. When the geodesic starting from the mean rotation in the manifold is mapped by the logarithm onto the tangent space at the mean, it is a straight line starting at the origin. The tangent space is called *exponential chart*. Hence, using the notations

$$r_2 \star r_1 = \log(\exp(r_2) \cdot \exp(r_1)), \quad r_1^{-1} = \log(\exp(r_1)^T)$$

for the rotation vectors r_1 and r_2 , the mean rotation \bar{r} satisfies

$$\sum_i (\bar{r}^{-1} \star r_i) = 0. \quad (13.22)$$

Weighting each rotation with π_i , yields the least squares problem:

$$\frac{1}{2} \sum_i \pi_i \|\bar{r}^{-1} \star r_i\|_2^2 \rightarrow \min. \quad (13.23)$$

The weighted mean can thus be estimated by

$$\hat{r}_{t+1} = \hat{r}_t \star \left(\frac{\sum_i \pi_i (\hat{r}_t^{-1} \star r_i)}{\sum_i \pi_i} \right). \quad (13.24)$$

The gradient descent method takes about 5 iterations until it converges.

The variance and the normal density on a Riemannian manifold can also be approximated, cf. [24]. Since, however, the variance is only used for diffusing the particles, a very accurate approximation is not needed. Hence, the variance of a set of rotations r_i is calculated in the Euclidean space \mathbb{R}^3 .

The twist representation used in [7, 26] and in chapters 11 and 12 is quite similar. Instead of a separation between the translation t and the rotation r , it describes a screw motion where the motion velocity θ also affects the translation. A twist $\hat{\xi} \in se(3)^3$ is represented by

$$\theta \hat{\xi} = \theta \begin{pmatrix} \hat{\omega} & v \\ 0 & 0 \end{pmatrix}, \quad (13.25)$$

where $\exp(\theta \hat{\xi})$ is a rigid body motion.

The logarithm of a rigid body motion $M \in SE(3)$ is the following transformation:

$$\theta \omega = \log(R), \quad v = A^{-1}t, \quad (13.26)$$

where

$$A = (I - \exp(\theta \hat{\omega}))\hat{\omega} + \omega \omega^T \theta \quad (13.27)$$

is obtained from the Rodriguez formula. This follows from the fact, that the two matrices which comprise A have mutually orthogonal null spaces when $\theta \neq 0$. Hence, $Av = 0 \Leftrightarrow v = 0$.

We remark that the two representations are identical for the joints where only a rotation around a known axis is performed. Furthermore, a linearization is not needed for the *ISA* in contrast to the pose estimation as described in chapters 11 and 12

³ $se(3)$ is the Lie algebra that corresponds to the Lie group $SE(3)$.

Pose prediction

The *ISA* can be combined with a pose prediction in two ways. When the dynamics are modelled by a Markov process for example, the particles of the current frame can be stored and predicted for the next frame according to the process as done in [12]. In this case, the *ISA* is already initialized by the predicted particles. But when the prediction is time consuming or when the history of previous poses is needed, only the estimate is predicted. The *ISA* is then initialized by diffusing the particles around the predicted estimate. The reinitialization of the particles is necessary for example when the prediction is based on local descriptors [13] or optical flow as discussed in chapter 11 and [5].

In our example, the pose is predicted by an autoregression that takes the global rigid body motions P_i of the last N frames into account [13]. For this purpose, we use a set of twists $\xi_i = \log(P_i P_{i-1}^{-1})$ representing the relative motions. By expressing the local rigid body motion as a screw action, the spatial velocity can be represented by the twist of the screw, see [20] for details.

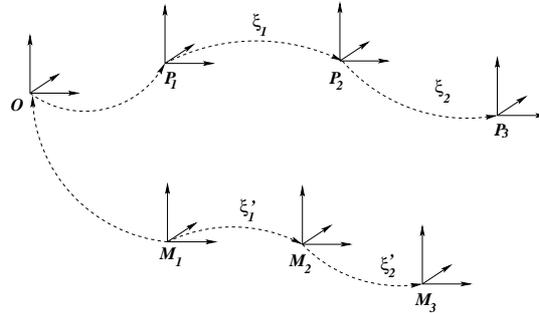


Fig. 13.10. Transformation of rigid body motions from prior data P_i in a current world coordinate system M_1 . A proper scaling of the twists results in a proper damping.

In order to generate a suited rigid body motion from the motion history, a screw motion needs to be represented with respect to another coordinate system. Let $\hat{\xi} \in se(3)$ be a twist given in a coordinate frame A . Then for any $G \in SE(3)$, which transforms a coordinate frame A to B , is $G\hat{\xi}G^{-1}$ a twist with the twist coordinates given in the coordinate frame B , see [20] for details. The mapping $\hat{\xi} \mapsto G\hat{\xi}G^{-1}$ is called the *adjoint transformation* associated with G .

Let $\xi_1 = \log(P_2 P_1^{-1})$ be the twist representing the relative motion from P_1 to P_2 . This transformation can be expressed as local transformation in the current coordinate system M_1 by the adjoint transformation associated with $G = M_1 P_1^{-1}$. The new twist is then given by $\hat{\xi}'_1 = G\hat{\xi}_1 G^{-1}$. The advantage of the twist representation is now, that the twists can be scaled by a factor $0 \leq \lambda_i \leq 1$ to damp the local rigid body motion, i.e. $\hat{\xi}'_i = G\lambda_i \hat{\xi}_i G^{-1}$. For given λ_i such that $\sum_i \lambda_i = 1$, the predicted pose is obtained by the rigid body transformation

$$\exp(\hat{\xi}'_N) \exp(\hat{\xi}'_{N-1}) \dots \exp(\hat{\xi}'_1). \tag{13.28}$$

Energy function

The energy function V of a particle x , which is used for our example, depends on the extracted silhouette and on some learned prior knowledge as in [12], but it is defined in a different way.

Silhouette: First of all, the silhouette is extracted from an image by a level set based segmentation as in [8, 27]. We state the energy functional E for convenience only and refer the reader to chapter 11 where the segmentation is described in detail. Let Ω^i be the image domain of view i and let $\Phi_0^i(\hat{x})$ be the contour of the predicted pose in Ω^i . In order to obtain the silhouettes for all r views, the energy functional $E(\hat{x}, \Phi^1, \dots, \Phi^r) = \sum_{i=1}^r E(\hat{x}, \Phi^i)$ is minimized, where

$$E(\hat{x}, \Phi^i) = - \int H(\Phi^i) \ln p_1^i + (1 - H(\Phi^i)) \ln p_2^i dx + \nu \int_{\Omega^i} |\nabla H(\Phi^i)| dx + \lambda \int_{\Omega^i} (\Phi^i - \Phi_0^i(\hat{x}))^2 dx. \quad (13.29)$$

In our experiments, we weighted the smoothness term with $\nu = 4$ and the shape prior with $\lambda = 0.04$.

After the segmentation, 3D-2D correspondences between the 3D model (X_i) and a 2D image (x_i) are established by the projected vertices of the 3D mesh that are part of the model contour and their closest points of the extracted contour that are determined by a combination of an iterated closest point algorithm [4] and an optic flow based shape registration [25]. More details about the shape matching are given in chapter 12. We write each correspondence as pair (X_i, x_i) of homogeneous coordinates.

Each image point x_i defines a projection ray that can be represented as Plücker line [20] determined by a unique vector n_i and a moment m_i such that $x \times n_i - m_i = 0$ for all x on the 3D line. Furthermore,

$$\|x \times n_i - m_i\|_2 \quad (13.30)$$

is the norm of the perpendicular error vector between the line and a point $x \in \mathbb{R}^3$. As we already mentioned, a joint j is represented by the rotation angle θ_j . Hence, we write $M(\omega, t)$ for the rigid body motion and $M(\theta_j)$ for the joints. Furthermore, we have to consider the kinematic chain of articulated objects. Let X_i be a point on the limb k_i whose position is influenced by s_i joints in a certain order. The inverse order of these joints is then given by the mapping ι_{k_i} , e.g., a point on the left shank is influenced by the left knee joint $\iota_{k_i}(4)$ and by the three joints of the left hip $\iota_{k_i}(3)$, $\iota_{k_i}(2)$, and $\iota_{k_i}(1)$.

Hence, the pose estimation consists of finding a pose x such that the error

$$err_S(x, i) := \left\| \left(M(\omega, t) M(\theta_{\iota_{k_i}(1)}) \dots M(\theta_{\iota_{k_i}(s_i)}) X_i \right)_{3 \times 1} \times n_i - m_i \right\|_2 \quad (13.31)$$

is minimal for all pairs, where $(\cdot)_{3 \times 1}$ denotes the transformation from homogeneous coordinates back to non-homogeneous coordinates.

Prior Knowledge: Using prior knowledge about the probability of a certain pose can stabilize the pose estimation as shown in [12] and [6]. The prior ensures that particles representing a familiar pose are favored in problematic situations, e.g., when the observed object is partially occluded. As discussed in chapter 11, the

probability of the various poses is learned from N training samples, where the density is estimated by a Parzen-Rosenblatt estimator [21, 29] with a Gaussian kernel

$$p_{pose}(x) = \frac{1}{(2\pi\sigma^2)^{d/2}N} \sum_{i=1}^N \exp\left(-\frac{\|x_i - x\|_2^2}{2\sigma^2}\right). \quad (13.32)$$

In our experiments, we chose the window size σ as the maximum second nearest neighbor distance between all training samples as in [12].

Incorporating the learned probability of the poses in the energy function has additional advantages. First, it already incorporates correlations between the parameters of a pose – and thus of a particle – yielding an energy function that is closer to the model and the observed object. Moreover, it can be regarded as a soft constraint that includes anatomical constraints, e.g. by the limited freedom of joints movement, and that prevents the estimates from self-intersections since unrealistic and impossible poses cannot be contained in the training data.

Altogether, the energy function V of a particle x is defined by

$$V(x) := \frac{1}{l} \sum_{i=1}^l err_S(x, i)^2 - \eta \ln(p_{pose}(x)), \quad (13.33)$$

where l is the number of correspondences. In our experiments, we set $\eta = 8$.

Results

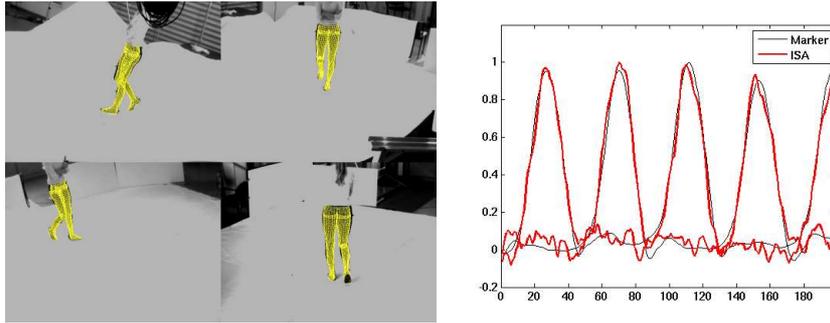


Fig. 13.11. **Left:** Results for a walking sequence captured by four cameras (200 frames). **Right:** The joint angles of the right and left knee in comparison with a marker based system.

In our experiments, we tracked the lower part of a human body using four calibrated and synchronized cameras. The walking sequence was simultaneously captured by a commercial marker based system⁴ allowing a quantitative error analysis. The training data used for learning p_{pose} consisted of 480 samples that were obtained from walking sequences. The data was captured by the commercial system before recording the test sequence that was not contained in the training data.

⁴ Motion Analysis system with 8 Falcon cameras.

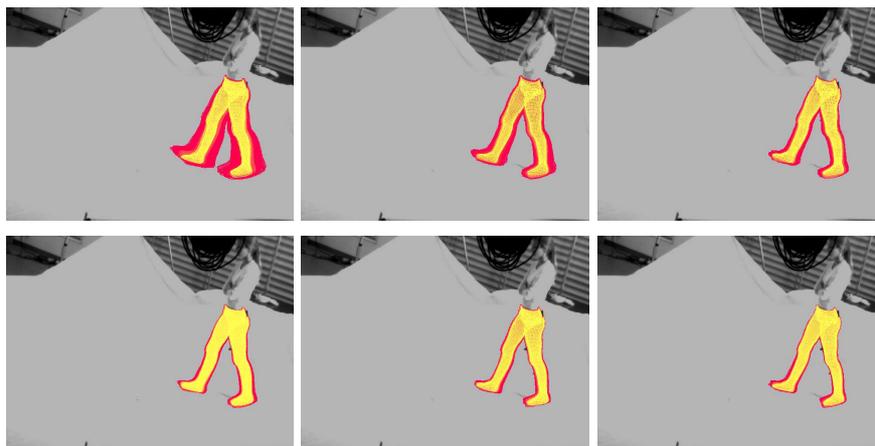


Fig. 13.12. Weighted particles at $t = 0, 1, 2, 4, 8,$ and 14 of *ISA*. Particles with a higher weight are brighter, particles with a lower weight are darker. The particles converge to the pose with the lowest energy as t increases.

The *ISA* performed well for the sequence consisting of 200 frames using a polynomial annealing scheme with $b = 0.7$, a dynamic variance scheme with $c = 0.3$, and the selection kernel (13.9). Results are given in Figure 13.11 where the diagram shows a comparison of the estimated knee-joint angles with the marker based system.

The convergence of the particles towards the pose with the lowest energy is illustrated for one frame in Figure 13.12. Moreover, it shows that variance of the particles decreases with an increasing number of annealing steps. This can also be seen from Figure 13.13 where the standard deviations for four parameters, which are scaled by c , are plotted. While the variances of the hip-joint and of the knee-joint decline rapidly, the variance of the ankle increases for the first steps before it decreases. This behavior results from the kinematic chain of the legs. Since the ankle is the last joint in the chain, the energy for a correct ankle is only low when also the previous joints of the chain are well estimated.

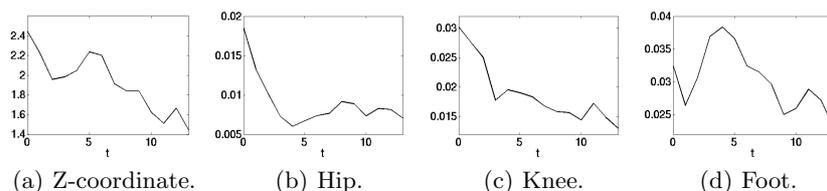


Fig. 13.13. Variance of the particles during *ISA*. The scaled standard deviations for the z-coordinate of the position and for three joint angles are given. The variances decrease with an increasing number of annealing steps.

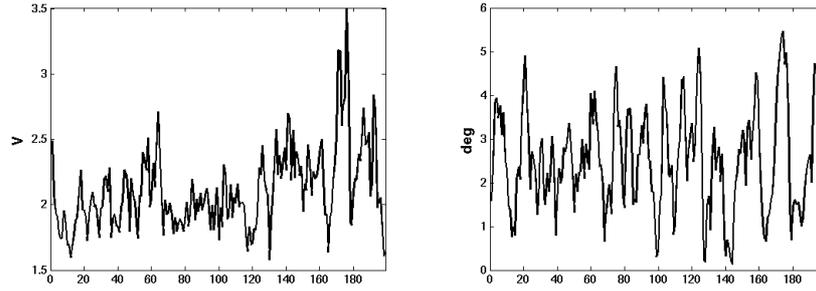


Fig. 13.14. **Left:** Energy of estimate for walking sequence (200 frames). **Right:** Error of estimate (left and right knee).

On the right hand side of Figure 13.14, the energy of the estimate during tracking is plotted. We also plotted the root-mean-square error of the estimated knee-angles for comparison where we used the results from the marker based system as ground truth with an accuracy of 3 degrees. For $n = 250$ and $T = 15$, we achieved an overall root-mean-square error of 2.74 degrees. The error was still below 3 degrees with 375 particles and $T = 7$, i.e. $n_T = 2625$. With this setting, the *ISA* took 7–8 seconds for approximately 3900 correspondences that were established in the 4 images of one frame. The whole system including segmentation, took 61 seconds for one frame. For comparison, the iterative method as used in chapter 12 took 59 seconds with an error of 2.4 degrees. However, we have to remark that for this sequence the iterative method performed very well. This becomes clear from the fact that no additional random starting points were needed. Nevertheless, it demonstrates that the *ISA* can keep up even in situations that are perfect for iterative methods.

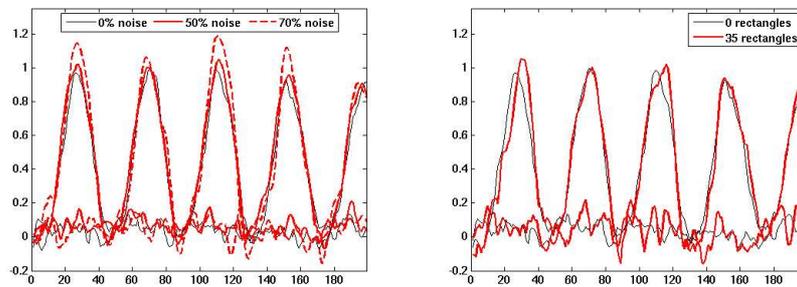


Fig. 13.15. **Left:** Random pixel noise. **Right:** Occlusions by random rectangles.

Figures 13.16 and 13.17 show the robustness in the presence of noise and occlusions. For the first sequence, each frame was independently distorted by 70% pixel noise, i.e., each pixel value was replaced with probability 0.7 by a value uniformly sampled

from the interval $[0, 255]$. The second sequence was distorted by occluding rectangles of random size, position, and gray value, where the edge lengths were in the range from 1 to 40. The knee angles are plotted in Figure 13.15. The root mean-square errors were 2.97 degrees, 4.51 degrees, and 5.21 degrees for 50% noise, 70% noise, and 35 occluding rectangles, respectively.

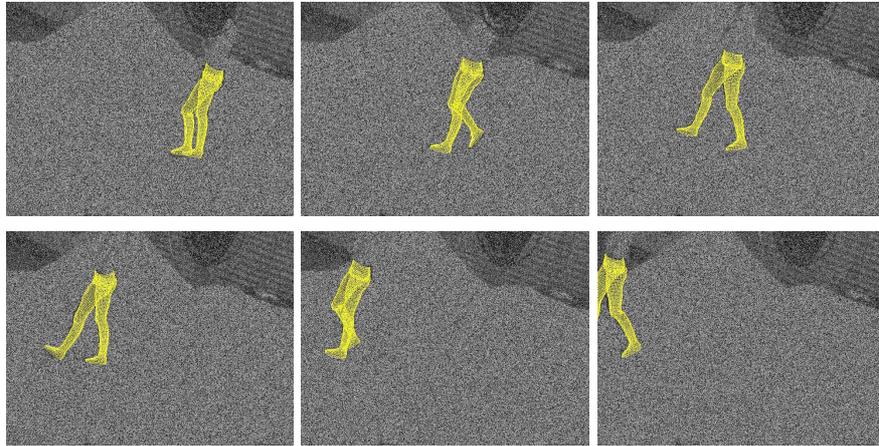


Fig. 13.16. Estimates for a sequence distorted by 70% random pixel noise. One view of frames 35, 65, 95, 125, 155, and 185 is shown.

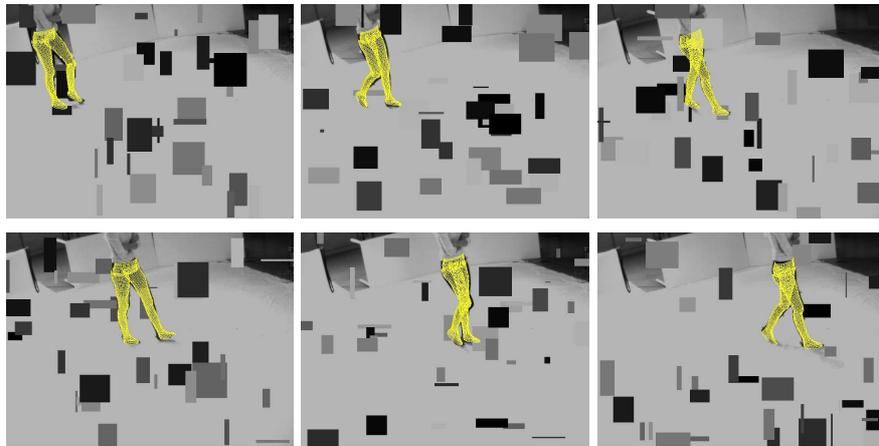


Fig. 13.17. Estimates for a sequence with occlusions by 35 rectangles with random size, color, and position. One view of frames 35, 65, 95, 125, 155, and 185 is shown.

13.5 Discussion

We introduced a novel approach for global optimization, termed interacting simulated annealing (*ISA*), that converges to the global optimum. It is based on an interacting particle system where the particles are weighted according to Boltzmann-Gibbs measures determined by an energy function and an increasing annealing scheme.

The variance of the particles provides a good measure of the confidence in the estimate. If the particles are all near the global optimum, the variance is low and only a low diffusion of the particles is required. The estimate, in contrast, is unreliable for particles with a high variance. This knowledge is integrated via dynamic variance schemes that focus the search on regions of interest depending on the confidence in the current estimate. The performance and the potential of *ISA* was demonstrated by means of two applications.

The first example showed that our approach can deal with local optima and solves the optimization problem well even for noisy measurements. However, we also provided some limitations of the dynamic variance schemes where standard global optimization methods might perform better. Since a comparison with other global optimization algorithm is out of the scope of this introduction, this will be done in future.

The application to multi-view human motion capturing, demonstrated the embedding of *ISA* into a complex system. The tracking system included silhouette extraction by a level-set method, a pose prediction by an auto-regression, and prior knowledge learned from training data. Providing an error analysis, we demonstrated the accuracy and the robustness of the system in the presence of noise and occlusions. Even though we considered only a relative simple walking sequence for demonstration, it already indicates the potential of *ISA* for human motion capturing. Indeed, a comparison with an iterative approach revealed that on the one hand global optimization methods cannot perform better than local optimization methods when local optima are not problematic as it is the case for the walking sequence, but on the other hand it also showed that the *ISA* can keep up with the iterative method. We expect therefore that the *ISA* performs better for faster movements, more complex motion patterns, and human models with higher degrees of freedom. In addition, the introduced implementation of the tracking system with *ISA* has one essential drawback for the performance. While the pose estimation is performed by a global optimization method, the segmentation is still susceptible to local minima since the energy function (13.29) is minimized by a local optimization approach.

As part of future work, we will integrate *ISA* into the segmentation process to overcome the local optima problem in the whole system. Furthermore, an evaluation and a comparison with an iterative method needs to be done with sequences of different kinds of human motions and also when the segmentation is independent of the pose estimation, e.g., as it is the case for background subtraction. Another improvement might be achieved by considering correlations between the parameters of the particles for the dynamic variance schemes, where an optimal trade-off between additional computation cost and increased accuracy needs to be found.

Acknowledgments

Our research is funded by the Max-Planck Center for Visual Computing and Communication. We thank Uwe Kersting for providing the walking sequence.

References

1. Ackley D. *A connectionist machine for genetic hillclimbing*. Kluwer, Boston, 1987.
2. Bäck T. and Schwefel H.-P. An overview of evolutionary algorithms for parameter optimization. *Evolutionary Computation*, 1:1–24, 1993.
3. Bauer H. *Probability Theory*. de Gruyter, Baton Rouge, 1996.
4. Besl P. and McKay N. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
5. Brox T., Rosenhahn B., Cremers D. and Seidel H.-P.. High accuracy optical flow serves 3-D pose tracking: exploiting contour and flow based constraints. In A.Leonardis, H.Bischof and A.Pinz, editors, *European Conference on Computer Vision (ECCV)*, LNCS 3952, pages 98–111. Springer, 2006.
6. Brox T., Rosenhahn B., Kersting U. and Cremers D.. Nonparametric Density Estimation for Human Pose Tracking. In *Pattern Recognition (DAGM)*. LNCS 4174, pages 546–555. Springer, 2006.
7. Bregler C., Malik J. and Pullen K. Twist Based Acquisition and Tracking of Animal and Human Kinematics. *International Journal of Computer Vision*, 56:179–194, 2004.
8. Brox T., Rosenhahn B. and Weickert J. Three-dimensional shape knowledge for joint image segmentation and pose estimation. In W.Kropatsch, R.Sablatnig, A.Hanbury, editors, *Pattern Recognition (DAGM)*, LNCS 3663, pages 109–116. Springer, 2005.
9. Deutscher J. and Reid I. Articulated body motion capture by stochastic search. *International Journal of Computer Vision*, 61(2):185–205, 2005.
10. Doucet A., deFreitas N. and Gordon N., editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer, New York, 2001.
11. Gall J., Potthoff J., Schnoerr C., Rosenhahn B. and Seidel H.-P. Interacting and Annealing Particle Systems – Mathematics and Recipes. *Journal of Mathematical Imaging and Vision*, 2007, To appear.
12. Gall J., Rosenhahn B. and Brox T. and Seidel H.-P. Learning for Multi-View 3D Tracking in the Context of Particle Filters. In *International Symposium on Visual Computing (ISVC)*, LNCS 4292, pages 59–69. Springer, 2006.
13. Gall J., Rosenhahn B. and Seidel H.-P. Robust Pose Estimation with 3D Textured Models. In *IEEE Pacific-Rim Symposium on Image and Video Technology (PSIVT)*, LNCS 4319, pages 84–95. Springer, 2006.
14. Geman S. and Geman D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.
15. Gidas B. *Topics in Contemporary Probability and Its Applications*, Chapter 7: Metropolis-type Monte Carlo Simulation Algorithms and Simulated Annealing, pages 159–232. Probability and Stochastics Series. CRC Press, Boca Raton, 1995.

16. Goldstein H. *Classical Mechanics*. Addison-Wesley, Reading, MA, second edition, 1980.
17. Grest D., Herzog D. and Koch R. Human Model Fitting from Monocular Posture Images. In G.Greiner, J.Hornegger, H.Niemann and M.Stamminger, editors, *Vision, Modeling and Visualization*. Akademische Verlagsgesellschaft Aka, 2005.
18. Kirkpatrick S., Gelatt C. Jr. and Vecchi M. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.
19. DelMoral P. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Probability and its Applications. Springer, New York, 2004.
20. Murray R.M., Li Z. and Sastry S.S. *Mathematical Introduction to Robotic Manipulation*. CRC Press, Baton Rouge, 1994.
21. Parzen E. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
22. Pennec X. and Ayache N. Uniform distribution, distance and expectation problems for geometric features processing. *Journal of Mathematical Imaging and Vision*, 9(1):49–67, 1998.
23. Pennec X. Computing the mean of geometric features: Application to the mean rotation. Rapport de Recherche RR-3371, INRIA, Sophia Antipolis, France, March 1998.
24. Pennec X. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.
25. Rosenhahn B., Brox T., Cremers D. and Seidel H.-P. A Comparison of Shape Matching Methods for Contour Based Pose Estimation. In R.Reulke, U.Eckhardt, B.Flach, U.Knauer and K.Polthier, editors, *11th International Workshop on Combinatorial Image Analysis (IWCIA), LNCS 4040*, pages 263–276. Springer, 2006.
26. Rosenhahn B., Brox T., Kersting U., Smith A., Gurney J. and Klette R. A system for marker-less human motion estimation. *Künstliche Intelligenz*, 1:45–51, 2006.
27. Rosenhahn B., Brox T. and Weickert J. Three-dimensional shape knowledge for joint image segmentation and pose tracking. In *International Journal of Computer Vision*, 2006, To appear.
28. Rosenhahn B., Perwass C. and Sommer G. Pose Estimation of Free-form Contours. *International Journal of Computer Vision*, 62(3):267–289, 2005.
29. Rosenblatt F. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27(3):832–837, 1956.
30. Szu H. and Hartley R. Fast simulated annealing. *Physic Letter A*, 122:157–162, 1987.
31. Theobalt C., Magnor M., Schueler P. and Seidel H.-P. Combining 2D Feature Tracking and Volume Reconstruction for Online Video-Based Human Motion Capture. In *10th Pacific Conference on Computer Graphics and Applications*, pages 96–103. IEEE Computer Society, 2002.
32. Tsallis C. and Stariolo D.A. Generalized simulated annealing. *Physica A*, 233:395–406, 1996.