Automatic Partitioning of High Dimensional Search Spaces associated with Articulated Body Motion Capture

Jonathan Deutscher
University of Oxford
Dept. of Engineering Science
Oxford, OX13PJ
United Kingdom
jdeutsch@robots.ox.ac.uk

Andrew Davison
University of Oxford
Dept. of Engineering Science
Oxford, OX13PJ
United Kingdom
ajd@robots.ox.ac.uk

Ian Reid
University of Oxford
Dept. of Engineering Science
Oxford, OX13PJ
United Kingdom
ian@robots.ox.ac.uk

Abstract

Particle filters have proven to be an effective tool for visual tracking in non-gaussian, cluttered environments. Conventional particle filters however do not scale to the problem of Human Motion Capture (HMC) because of the large number of degrees of freedom involved.

Annealed Particle Filtering (APF), introduced by Deutscher et al [3], tackled this by layering the search space and was shown to be a very effective tool for HMC.

In this paper we improve upon and extend the APF in two ways. First we develop a hierarchical search strategy which automatically partitions the search space without any explicit representation of the partitions. Then we introduce a crossover operator (similar to that found in Genetic Algorithms) which improves the ability of the tracker to search different partitions in parallel.

We present results for a simple example to demonstrate the new algorithm's implementation and then apply it to the considerably more complex problem of Human Motion Capture with 34 degrees of freedom.

1. Introduction

Marker-based human motion capture has been used commercially [14] for a number of years with applications found in special effects and biometrics. The use of markers however is intrusive, necessitates the use of expensive specialised hardware and can only be used on footage taken especially for that purpose. A markerless system of human motion capture could be run using conventional cameras and without the use of special apparel or other equipment. Combined with today's powerful off-the-shelf PC's, cost-effective and real-time markerless human motion capture has for the first time become a possibility. Such a system would have a greater number of applications than its marker-based predecessor ranging from intelligent surveillance to character animation and computer interfacing.

Past research into HMC has concentrated on the articulated-model based approach. The reason this approach is popular is the high level output it produces in the form of a model configuration for each frame. This output can easily be used by higher-order processes to perform tasks such as character animation and action recognition.

The main challenge in articulated body motion tracking is the large number of degrees of freedom (at least 30 for a realistic human model) to be recovered. Search algorithms, either deterministic or stochastic, that search such a space without constraint, fall foul of exponential computational complexity.

One way researchers have tackled this problem is to relax constraints arising from articulation, and track limbs as if their motions were independent. This assumes that the different parts of the body can in fact be tracked independently, a feat usually requiring special markers or apparel.

Another popular approach is to introduce constraints — either labelling using markers or colour coding, prior assumptions about motion trajectories or view restrictions. It is also possible to restrict the range of movement of the subject. This approach has been pursued by Hogg [6], Rohr [11] and Niyogi [9]. All three assume the subject is walking (usually in a straight line). Rohr even reduces the dimension of the problem to the phase of the walking cycle. Goncalves [5] and Deutscher [2] assume a constant angle of view of the subject as does Bregler [1] and Rehg [10]. Such an approach greatly restricts the resulting tracker's generality.

Using a very strong dynamical model is another way to introduce constraints. Sidenbladh *et al* [12] have taken a mathematically rigorous approach to full-body tracking based on Condensation [7] where they employ learned dynamical models and a generative model of image formation. They tracked short sequences of 3D motion from a single camera, though the very strong dynamical models used restrict the applicability of the system to a narrow range of movement and the system runs slowly due to the large number of particles required.

Others have proposed various forms of search space de-

ISBN 0-7695-1272-0/01 $10.00 (C) 2001 IEEE

composition as a solution, most notably Gavrila *et al* [4]. MacCormick *et al's* [8] work on Partitioned Sampling and Sullivan *et al's* [13] Layered Sampling are versions of the Condensation algorithm that combine search space partitioning or layering with a particle filter framework. Although reducing the size of search space dramatically these approaches either require restrictive assumptions about pose, subject and environment, or are still too inefficient to be employed for general HMC.

Annealed Particle Filtering (APF) [3] has been proposed as a general and robust solution for search in non-gaussian, high-dimensional spaces with specific application to Human Motion Capture (HMC). It uses a continuation principle, based on annealing, to introduce the influence of narrow peaks in the fitness function gradually, and has been shown to track complex human movement without the use of extra constraints such as labelled markers, pose assumptions, restricted movement or colour coding.

In this paper we improve upon and extend the APF in two ways. We develop a hierarchical search strategy which *automatically* partitions the search space without any explicit representation of the partitions. Then we introduce a crossover operator (similar to that found in Genetic Algorithms) which improves the ability of the tracker to search different partitions in parallel. Both these measures, brought together in the Partitioned Annealed Particle Filter (PAPF), increase the speed of the original APF by a factor of 4.

We present results for the simple example of tracking a planar articulated arm to demonstrate the new algorithm's implementation and then apply it to the considerably more complex problem of Human Motion Capture with 34 degrees of freedom.

2. Search Space Decomposition

2.1 Conventional Search Space Decomposition

A very effective method of reducing the computational effort required to search a high DOF space is *Search Space Decomposition*. If one section of a search space can be searched independently then it can be used as a constraint for reducing the rest of the search space. In the context of HMC for example one could try to match just the torso and head to an image, and once these sections have been localised we could progress to the arms and legs.

A successful partitioning of the search space in this way reduces the cost of searching the space to one that increases linearly with the number of partitions instead of one that increases exponentially with the number of DOF. This would obviously be of great computational benefit when matching high DOF models like those used in HMC to model humans.

However, there are a number of problems with using simple Search Space Decomposition for realistic HMC.

- Each partition requires the formulation of a separate weighting function [1]. It is not always possible to decompose the overall weighting function into separate weighting functions for each partition and even if this can be done it is difficult to ensure their reliability, as previously mentioned.
- Even before the separate weighting functions are devised, decisions have to be made about how many partitions to have and where to divide the search space into these separate partitions. This is usually done in an intuitive fashion about what will "work best", ie. separating the limbs into separate partitions and the torso and head into another.

  The problem with this particular approach is that it is very difficult to find a good match for just the torso and head, especially in rotation about the vertical axis, without using information from the arms and legs at the same time (see figure 1). In fact it is very hard to reliably match any section of a human model to an image independently of the other sections without the help of markers, color cues or pose assumptions. The best partition points may even change during tracking. It would be best not to explicitly make these decisions at all.

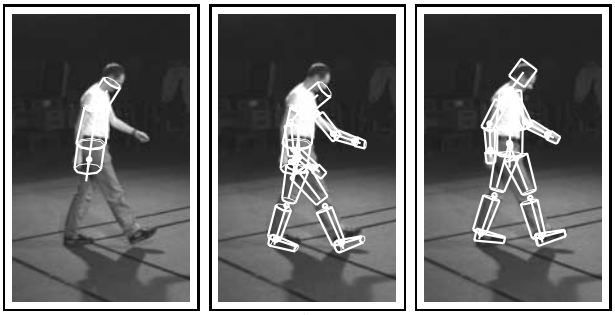

Figure 1: **Hierarchical matching of model partitions.** Matching just the head and torso of a model to the image results in a poor fit (a). If this result is used to constrain the search for the limbs as is Hierarchical Search Space Decomposition we are left with an unsatisfactory outcome as seen in (b) illustrating the dependence of the partitions. Searching over the entire unpartitioned space results in a good match as seen in (c). The weighting function was a sum of squared difference function of edge and foreground information. Information from three cameras was used but only one view shown for clarity.

2.2 Search Space Decomposition using predicted values

Gavrila [4] avoids the problem of constructing multiple weighting functions by predicting configurations for each partition.

The model is again split into torso-head and limb partitions (decisions about where to partition the search space

[1] Defined here as the function used to determine the degree of fit between a model (or one of its partition's) configuration $\mathbf{X}$ and a given image $\mathbf{Z}$, denoted as $w(\mathbf{X}, \mathbf{Z})$.

ISBN 0-7695-1272-0/01 $10.00 (C) 2001 IEEE

still have to be made before tracking). Then the goodness of fit for the first partition is optimised over the first partition parameters, keeping the other parameters constant at the values predicted by a dynamical model. Having found the best configuration for the torso-head partition he is then able to search for the best configuration for the remaining limb partitions — again without using separate weighting functions.

Although Gavrila has presented some excellent results using this method he acknowledges that using an unoccluded subject wearing tight color-coded clothing produces "a well behaved similarity measure derived from multiple views . . . [which] is likely to lead a search landscape with fairly wide and pronounced maxima around the correct parameter values" [4].

When applied in other circumstances in which the topology of the matching function is not so well behaved, tracking often fails catastrophically. Consider the simple articulated arm, tracked using a sum of squared difference (SSD) measure between image and model (the exact details are unimportant) and split into four partitions.

When tracking this arm using Gavrila's predictive SSD (see figure 3) even a small error in the predicted model configuration generates an error in the final model configuration at each frame. This error in turn produces a larger error in the predicted configuration for the next frame which produces a larger error in the calculated value and so on an so forth until tracking fails.

This shows that the separate partitions of an articulated object cannot be considered independent (even if they are nearly so) even near the maximum of the weighting function — which is where the predicted values should place us. Our experiments show that this characteristic of articulated bodies is even more prevalent in the much more complex scenario of full-body HMC.

2.3 Search Space Decomposition using Particle Filters

Ideally we require a method in which the near independence of each partition can be exploited without making the catastrophic assumption of complete independence. The use of particle filters can provide a framework for elegant solutions and a number have emerged in recent years.

Variations on the Condensation [7] algorithm have been proposed (or could conceivably be used) for tracking articulated objects, in particular MacCormick *et al's* [8] work on Partitioned Sampling and Sullivan *et al's* [13] Layered Sampling.

These approaches do not require each partition to be completely independent, nor do they require completely reliable partition weighting functions. However they still present a number of other problems:

1. Pre-determined partitions of the search space are still necessary.
2. Weighting functions are still needed for each partition.

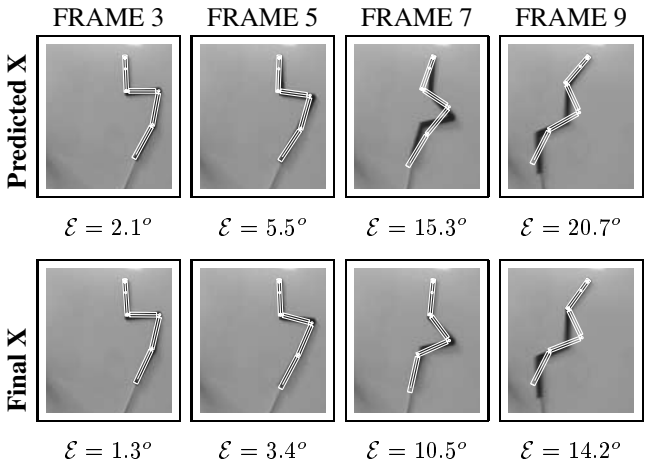

Figure 2: **Tracking a simple articulated arm using Predictive SSD.** The configuration of the arm, $\mathbf{X} = (x_1, x_2, x_3, x_4)$ (seen in figure 3) is split into 4 partitions, one for each parameter.
The top row shows the predicted configuration — computed using a constant position model of motion — superimposed on the actual articulated arm for each frame. The bottom row shows the final configuration output by the tracker for each frame, found using the the Predictive SSD framework.
The error values $\mathcal{E}$ are the Euclidean distance to the true configuration which was found using exhaustive search over the entire space.

3. Neither method increases efficiency enough (a factor 5-10 times is quoted by MacCormick [8]) to enable tracking though very high DOF spaces.

By contrast the Annealed Particle Filter, also based on a particle filter framework, is much more efficient than either of these methods and has already been proven to be effective for HMC.

3. Annealed Particle Filter

The Annealed Particle Filter (APF) (Deutscher *et al* [3]) has proven to be a very robust and simple way to search high-dimensional search spaces efficiently, and has been demonstrated on sequences of complex Human Motion. In the remainder of this paper we first briefly summarise the APF algorithm and then show how it can be improved and extended to produce a tracking framework that does not require the specification of partitions or multiple weighting functions and automatically takes advantage of any independence that may exist between different regions of the search space to increase tracking efficiency without sacrificing tracking accuracy.

The APF algorithm is briefly outlined below, we recommend that the diligent readers peruse Deutscher *et al's* [3] paper for the full details.

3.1 The APF algorithm

A series of weighting functions $w_0(\mathbf{Z}, \mathbf{X})$ to $w_M(\mathbf{Z}, \mathbf{X})$ are employed which measure the goodness-of-fit of hypotheti-

ISBN 0-7695-1272-0/01 $10.00 (C) 2001 IEEE

cal pose $\mathbf{X}$ with respect to image data $\mathbf{Z}$. Each weighting function w_m differs only slightly from w_{m-1}. The function w_M is designed to be very broad, representing the overall trend of the search space while w_0 should be very peaked, emphasising local features. This is achieved by setting

$$w_m(\mathbf{Z}, \mathbf{X}) = w(\mathbf{Z}, \mathbf{X})^{\beta_m}, \tag{1}$$

for $\beta_0 > \beta_1 > \ldots > \beta_M$, where $w(\mathbf{Z}, \mathbf{X})$ is the original weighting function. Because it is not the aim to sample from $w(\mathbf{Z}, \mathbf{X})$, but only to find its maximum it is not required that $\beta_0 = 1$.

One *annealing run* consisting of several iterations of a single algorithm (or layers) is performed at each time t_k using image observations $\mathbf{Z}_k$. The state of the tracker after each layer m of an annealing run is represented by a set of N weighted particles

$$\mathcal{S}^{\pi}_{k,m} = \{(\mathbf{s}^{(0)}_{k,m}, \pi^{(0)}_{k,m}) \ldots (\mathbf{s}^{(N)}_{k,m}, \pi^{(N)}_{k,m})\}. \tag{2}$$

Each particle $(\mathbf{s}^{(i)}_{k,m}, \pi^{(i)}_{k,m})$ contains an instance $\mathbf{s}^{(i)}_{k,m}$ of the multi-variate model configuration $\mathbf{X}$, and a corresponding particle weighting $\pi^{(i)}_{k,m}$. An unweighted set of particles will be denoted

$$\mathcal{S}_{k,m} = \{(\mathbf{s}^{(0)}_{k,m}) \ldots (\mathbf{s}^{(N)}_{k,m})\}. \tag{3}$$

Each annealing run can be broken down as follows.

1. For every time step t_k an annealing run is started at layer M, with $m = M$.
2. Each layer of an annealing run is initialised by a set of un-weighted particles $\mathcal{S}_{k,m}$.
3. Each of these particles is then assigned a weight

   $$\pi^{(i)}_{k,m} \propto w_m(\mathbf{Z}_k, \mathbf{s}^{(i)}_{k,m}) \tag{4}$$

   which are normalised so that $\sum_N \pi^{(i)}_{k,m} = 1$. The set of weighted particles $\mathcal{S}^{\pi}_{k,m}$ has now been formed.
4. N particles are drawn randomly from $\mathcal{S}^{\pi}_{k,m}$ with replacement and with a probability equal to their weighting $\pi^{(i)}_{k,m}$. As the n^{th} particle $\mathbf{s}^{(n)}_{k,m}$ is chosen it is used to produce the particle $\mathbf{s}^{(n)}_{k,m-1}$ using

   $$\mathbf{s}^{(n)}_{k,m-1} = \mathbf{s}^{(n)}_{k,m} + \mathbf{B}_m \tag{5}$$

   where $\mathbf{B}_m$ is a multi-variate gaussian random variable with covariance matrix $\mathbf{P}_m$ and mean $\mathbf{0}$.
5. The set $\mathcal{S}_{k,m-1}$ has now been produced which can be used to initialise layer $m-1$. The process is repeated until we arrive at the set $\mathcal{S}^{\pi}_{k,0}$.
6. $\mathcal{S}^{\pi}_{k,0}$ is used to estimate the optimal model configuration $\mathcal{X}_k$ using

   $$\mathcal{X}_k = \sum_{i=1}^{N} \mathbf{s}^{(i)}_{k,0} \pi^{(i)}_{k,0}. \tag{6}$$

7. The set $\mathcal{S}_{k+1,M}$ is then produced from $\mathcal{S}^{\pi}_{k,0}$ using

   $$\mathbf{s}^{(n)}_{k+1,M} = \mathbf{s}^{(n)}_{k,0} + \mathbf{B}_0. \tag{7}$$

   This set is then used to initialise layer M of the next annealing run at t_{k+1}.

If a dynamical model of motion was introduced (the current system could be considered to employ a constant position model) then $\mathbf{P}_0$ would be the process noise covariance. In practice $\mathbf{P}_0$ is restricted to a diagonal matrix where each element is allocated a value equal to half the maximum expected movement of the corresponding model configuration parameter over one time step, plus a component to allow for gross tracking errors. In this way particles in the set $\mathcal{S}_{k+1,M}$ should cover all possible movements of the subject between time t_k and t_{k+1}. The amount of diffusion added to each successive annealing layer should decrease at the same rate as the resolution of the set $\mathcal{S}_{k,m}$ increases. It is suggested by Deutscher *et al* that setting

$$\mathbf{P}_m = \mathbf{P}_0(\alpha_M \alpha_{M-1} \ldots \alpha_m) \tag{8}$$

produces good results.

As it stands Annealed Particle Filtering is (one of) the most efficient, and general methods available for robust Human Motion Capture. In this paper a method is proposed which combines the robust nature of APF with the efficiencies gained by partitioning the search space to dramatically increase the speed and reliability of tracking.

4. Hierarchical partitioning theory

We now incorporate *Search Space Decomposition* into the framework of the Annealed Particle Filter, initially demonstrating the concept on a cut-down example.

Consider the simple task of tracking an articulated arm as seen in figure 3. The arm consists of four segments, each joined by a swivelling joint with one end rooted on the spot. A configuration of the arm is described by an instance of the state variable $\mathbf{X} = (x_1, x_2, x_3, x_4)$. The weighting function $w(\mathbf{Z}, \mathbf{X})$ required for the APF is computed by a Sum of Squared Difference (SSD) measure between a model template the image, the exact details are not important.

The set $\mathcal{S}_{k,m}$ is initialised with particles uniformly distributed over a range of $\mathbf{X}$ that we know to contain the actual position of the arm. This results in a large and similar variance for each parameter of $\mathbf{X}$ over all the particles in $\mathcal{S}_{k,m}$ as can be seen in figure 4a. After calculating a weight $\pi^{(i)}_{k,m}$ for each particle using $w_m(\mathbf{Z}_k, \mathbf{s}^{(i)}_{k,m})$ we then proceed to step 4 of the APF and draw N particles from $\mathcal{S}_{k,m}$ with replacement and probability proportional to each particle's weight.

Consider the set $\mathcal{S}_{k,m}$ so produced before the addition of any noise. In a typical annealing run the individual parameters of each particle were found to have variance as detailed in figure 4b. Note here that the variance of x_1 has

ISBN 0-7695-1272-0/01 $10.00 (C) 2001 IEEE

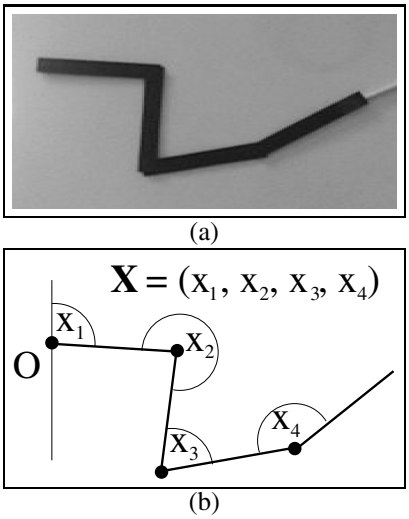

Figure 3: **A planar articulated arm** with 4 DOF is shown (a). It consits of four links connected by swivelling joints and rooted at **O**. The configuration of the arm is described by $\mathbf{X} = (x_1, x_2, x_3, x_4)$ as seen in (b).

been greatly reduced while the other parameters x_2, x_3 and x_4 have been hardly reduced at all. The variance of any parameter can be considered (with a number of acceptable caveats) to be directly related to the degree to which the maximal value for that parameter has been determined. Figure 4b shows that x_1 has been localised down to a very small area of its range simply because it dominates the topology of the search space whereas each particle's values for x_2, x_3 and x_4 had very little influence on whether it was selected or not. In effect we see here an automatic partitioning of the state space into soft partitions according each parameter's topological dominance. This should always occur as long as N is not high enough to allow saturation of the range of $\mathbf{X}$.

The problem with classical APF arises with the addition of noise to each particle upon selection. According to equations 5 and 8 an equal amount of noise should be added to each particle. This results in a parameter variance profile like that seen in figure 4c with the localisation of x_1 seen in figure 4b all but wiped out by the excessive addition of noise.

If instead the amount of noise added to each parameter $x_1 \ldots x_4$ of each selected particle is proportional to the variance of that parameter over the set of particles these gains will be protected from disruption. Instead we will arrive at the situation seen in figure 4(f) where enough noise has been added to keep the APF working, but not enough to wipe out any localisation gains.

If this method of determining $\mathbf{P}_i$ is continued through all the annealing layers we can see that each parameter is localised in turn, with some degree of overlap as seen in figure 5. This is exactly the kind of hierarchical soft partitioning that was desired and no explicit partition boundaries or functions were required.

A good measure of the performance of a particle filter

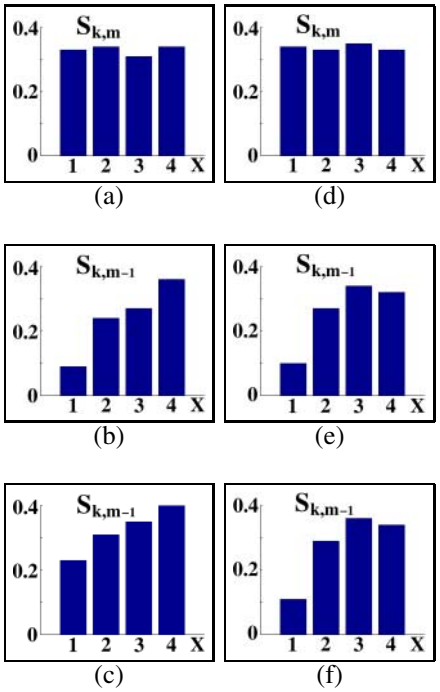

Figure 4: **Graphs of parameter variance illustrate the failure of conventional APF.** On the left graphs a, b and c plot the variance of each parameter of $\mathbf{X} = (x_1, x_2, x_3, x_4)$ through the first annealing run of the APF when tracking the articulated arm seen in figure 3. Graphs d, e and f show the same information for the improved APF as described in section 4.1. Graphs a and d show the variances of the initial set $\mathcal{S}^{\pi}_{k,m}$, displaying equal variances for each parameter. Graphs b and e show the variances of the set $\mathcal{S}^{\pi}_{k,m-1}$ before the addition of noise. Note that in both b and e, x_1 has a very small variance indicating advanced localisation, however x_2, x_3 and x_4 have been reduced only a little. up to this point the algorithms are the same and any differences between b and e are random. After the addition of noise in the original APF the localisation of x_1 has been greatly degraded as seen in graph c, however when noise is added in proportion to each parameter's variance the localisation of x_1 is preserved as seen in graph f.

is the number of particles needed to successfully track an object, a good indication of the computational resources required to run the tracker. It was found that this new technique provided a 2 fold increase in tracking performance according to this measure.

4.1 The amended APF algorithm

The changes to the APF can be formalised as follows. Step 4 of the APF algorithm described in section 3 is amended so that at layer m, $\mathbf{P}_m$ is set to be proportional to the variance of the particles in $\mathcal{S}_{k,m}$ as it exists before the addition of noise, ie.

$$\mathbf{P}_m \propto \frac{1}{N}\sum_{i=1}^{N}(\mathbf{s}^{(i)}_{k,m} - \mathbf{s}^{av}_{k,m}).(\mathbf{s}^{(i)}_{k,m} - \mathbf{s}^{av}_{k,m})^T. \qquad (9)$$

ISBN 0-7695-1272-0/01 $10.00 (C) 2001 IEEE

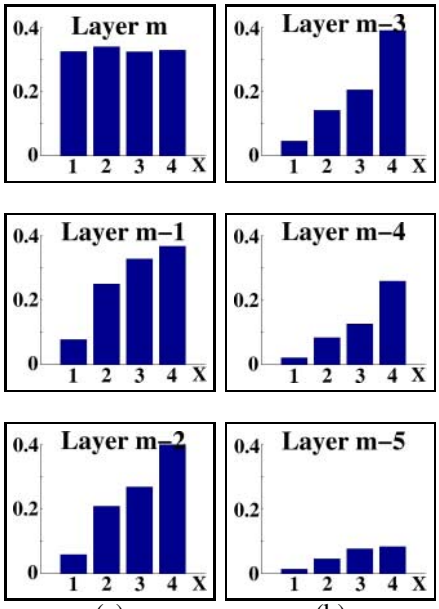

Figure 5: **Variance reduction in improved APF.** Here we see the orderly reduction of each of the four parameter's variances from most dominant (x_1) to least dominant (x_4) over 6 layers of the annealing process. Using the improved APF results in a 2-fold increase in efficiency over the classical APF. Tracker efficiency was measured by the minimum number of particles needed to successfully track the articulated arm over 40 frames.

5. Parallel partitioning theory

5.1 The need for parallel partitions

Consider the articulated object found in figure 6 which consists of two articulated arms joined at a stationary hinge. This configuration is a much simplified version of that found in HMC when using a model with arms and legs.

The soft hierarchical partitioning described in section 4 will provide some increase in efficiency over conventional APF when applied to tracking this assembly, localising x_1 and x_4 together, then x_2 and x_5 and finally x_3 and x_6. However if we were to decouple the search space and localise each arm independently the computational effort required for tracking would be reduced considerably.

One possibility, of course, would be to introduce a hard partition between the two arms and conduct two separate searches. However, as in the previous section we seek to avoid commitment to specific partitions since in more complex examples the optimal partitioning may not be obvious and it may indeed change over time as the degree of interaction between different parts of a model changes — such as when the legs cross during walking.

5.2 The crossover operator

Many people comment on the similarity between Particle Filters and Genetic Algorithms. Both employ a [set — population] of [particles — individuals] coded by a [state vector — genetic sequence] from which the best [particles — individuals] are chosen to be propagated to the next [time-step — generation] in the hope of finding the [maximum of some function — fittest possible individual].

One glaring difference between GA's and a typical particle filter is the lack of a crossover operator in the particle filter which in conventional GA is meant to simulate the breeding of individuals and the sharing of genetic information.

The reason touted by GA enthusiasts for the crossover operator is to enable the GA to search different parts of the parameter space (partitions) in parallel. In fact the use of the crossover operator encourages the survival of short, highly fit sections of the parameter space known in some GA literature as building blocks. This is done in the hope that when highly fit building blocks are brought together they will have a good chance of forming a very fit complete individual. It would be good to note that these building blocks are effectively optimised in parallel without any specification of their boundaries or appropriate building block (partition) weighting functions, exactly the kind of behaviour we are looking for.

We now describe how to incorporate the crossover operator into the framework of the APF and examine the effect via a simple example.

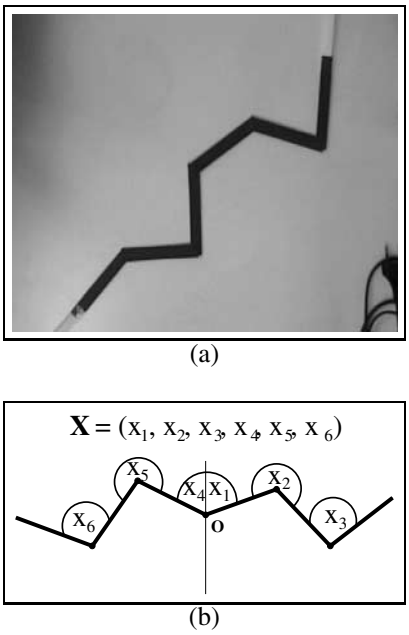

Figure 6: **Two planar articulated arms** consisting of 3 segments each and both rooted to point **O** (as seen in b) are used to demonstrate the effectiveness of the crossover operator. The configuration of the arms is described by $\mathbf{X} = (x_1, \ldots, x_6)$ as seen in (b).

ISBN 0-7695-1272-0/01 $10.00 (C) 2001 IEEE

5.3 Inclusion of the crossover operator in the APF

The inclusion of the crossover operator can be formalised as follows. In step 4 of the APF (as described in section 3) at annealing layer m, the i^{th} particle of $\mathcal{S}_{k,m-1}$ is created by drawing two particles from $\mathcal{S}^{\pi}_{k,m}$ with probability proportional to their respective weights. Two parameter indices γ and ϵ are chosen randomly and the two selected particles $\mathbf{s}^{(a)}_{k,m} = (x^a_1 \dots x^a_L)$ and $\mathbf{s}^{(b)}_{k,m} = (x^b_1 \dots x^b_L)$ are combined to form the new particle $\mathbf{s}^{(i)}_{k,m-1}$ where

$$\mathbf{s}^{(i)}_{k,m-1} = \left(x^a_1, \dots, x^a_\gamma, x^b_{\gamma+1}, \dots x^b_\epsilon, x^a_{\epsilon+1}, \dots, x^a_L\right). \tag{10}$$

Noise is then added to each particle as detailed in section 4.1.

5.4 Testing the crossover operator

To determine if there is any benefit to the crossover operator two articulated objects were tracked. The first seen in figure 3, was the experiment from section 4, an un-branched articulated arm. The second as seen in figure 6 is two articulated arms rooted to the same position.

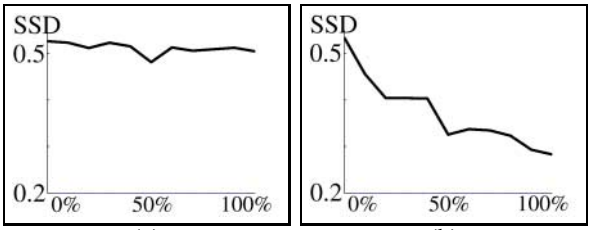

Figure 7: **The crossover operator in action.** The Sum of Squared Difference (SSD) match between model and image obtained after a set number of annealing layers is plotted against the percentage of particles generated using the crossover operator at each annealing layer. Graph (a) shows the result for the articulated arm seen in figure 3 where no benefit to using the crossover operator is seen although importantly no degradation in performance is seen either. Graph (b) shows the result for the articulated arms seen in figure 6 where a steady improvement in tracking performance is seen when increasing the percentage of crossover. This shows that the crossover operator is able to decouple sections of the search space and enables the APF to search them in parallel improving tracker performance.

As we can see in figure 7, the object consisting of branched arms was more effectively localised by the APF that employed the crossover operator whereas there was no difference when it was applied to the non-branched object. This gives a good graphical illustration of what the crossover operator is actually doing, ie partitioning sections of the search space which can be tracked in parallel. A good indication of the increased speed provided by the crossover operator when tracking branched objects is again the number of particles needed for successfull tracking. This number was which was reduced by a factor of 2 with the introduction of the crossover operator.

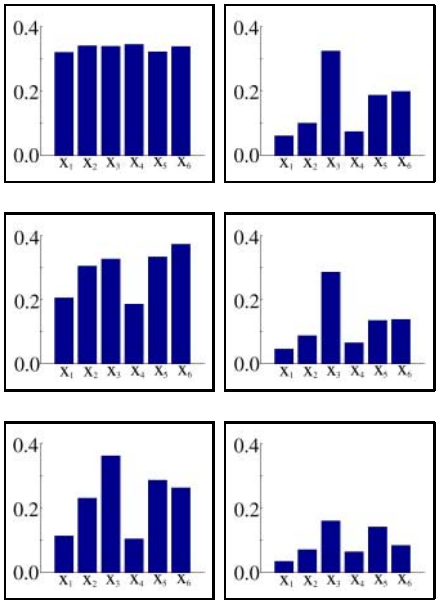

Figure 8: **Variance reduction for the parallel arms.** When the APF with crossover operator is applied to the articulated arms seen in figure 6 we get the pattern of variance reduction seen above. The graphs show the parameters describing each arm (x_1, x_2, x_3 and x_4, x_5, x_6) being localised in order of decreasing topological dominance in line with what would be intuitively expected.

6. Full body tracking

The new Partitioned Annealed Particle Filter (the PAPF, which incorporates all the improvements to the APF introduced in this paper) was tested on a number of challenging sequences of human movement including walking with turns (figure 9), running around in a random fashion (figure 10) and handstands (figure 11). Refer to our website "www.robots.ox.ac.uk/˜jdeutsch/HMC/" for video and more images. The weighting function used was a sum of squared difference function of edge and foreground information. Tests indicated that tracking performance, again measured by the number of particles needed to successfully track a sequence, was improved by a factor of 4 when comparing the new PAPF to the original APF. As a result the PAPF required on average 15 seconds to process one frame whereas APF required around 60 seconds when run on a single processor 1GHz pIII Linux box. The potential for threaded computing, obviously applicable to particle filters has not yet been explored.

7. Conclusion

The main challenge in articulated body motion tracking is the large number of degrees of freedom (around 30) to be recovered. Search algorithms, either deterministic or stochas-

ISBN 0-7695-1272-0/01 $10.00 (C) 2001 IEEE

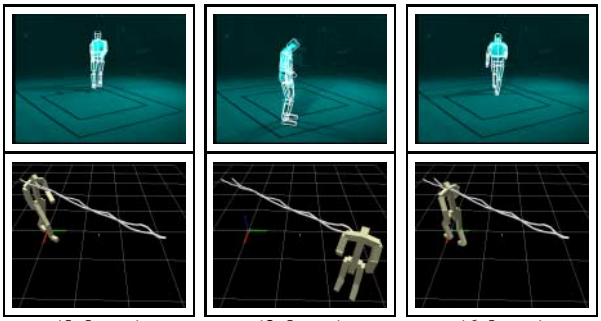

Figure 9: **Tracking a walking person.**

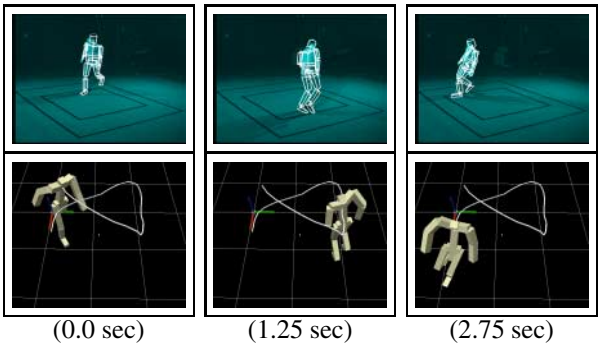

Figure 10: **Tracking a running person.**

tic, that search such a space without constraint, fall foul of exponential computational complexity.

Annealed Particle Filtering (APF) [3] has been proposed as a general and robust solution to this problem and has been shown to track complex human movement without the use of extra constraints such as labelled markers or colour coding.

Others have proposed various forms of search space decomposition as a solution. Although reducing the size of search space dramatically these approaches usually require restrictive assumptions and tailored scenarios that cannot be tolerated in a general and robust tracker.

This paper proposed an improvement of the APF algorithm which automatically incorporates many of advantages of hierarchical and parallel partitioning of the search space

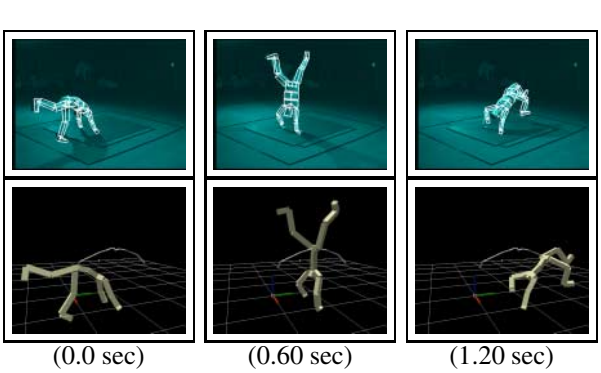

Figure 11: **Tracking a person performing a handstand.**

while retaining the generality and robustness fo the original APF algorithm. The new algorithm termed Partitioned Annealed Particle Filtering (PAPF) was demonstrated on numerous challenging sequences of Human Motion, providing a huge gain in efficiency over the already swift APF, bringing closer the prospect of real-time, robust and generally applicable Human Motion Capture.

Acknowledgements This work was supported by Vicon Motion Systems (www.vicon.com) and EPSRC grant GR/M15262. We are also grateful for discussions with Andrew Blake.

References

[1] Bregler, C., and Malik, J. Tracking people with twists and exponential maps. In *Proc. CVPR* (1998).

[2] Deutscher, J., Blake, A., North, B., and Bascle, B. Tracking through singularities and discontinuities by random sampling. In *Proc. 7th Int. Conf. on Computer Vision* (1999), vol. 2, 1144–1149.

[3] Deutscher, J., Blake, A., and Reid, I. Articulated body motion capture by annealed particle filtering. In *Proc. Conf. Computer Vision and Pattern Recognition* (2000), vol. 2, 1144–1149.

[4] Gavrila, D., and Davis, L. 3d model-based tracking of humans in action: a multi-view approach. *Proc. Conf. Computer Vision and Pattern Recognition* (1996), 73–80.

[5] Goncalves, L., di Bernardo, E., Ursella, E., and Perona, P. Monocular tracking of the human arm in 3D. In *Proc. 5th Int. Conf. on Computer Vision* (1995), 764–770.

[6] Hogg, D. Model-based vision: a program to see a walking person. *J. Image and Vision Computing 1*, 1 (1983), 5–20.

[7] Isard, M., and Blake, A. Visual tracking by stochastic propagation of conditional density. In *Proc. 4th European Conf. Computer Vision* (Cambridge, England, Apr 1996), 343–356.

[8] MacCormick, J., and Blake, A. Partitioned sampling, articulated objects and interface-quality hand tracking. In *Accepted to ECCV 2000* (2000).

[9] Niyogi, S., and Adelson, E. Analysing and recognising walking figures in xyt. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition* (1994), 469–474.

[10] Rehg, J., and Morris, D. Singularities in articulated object tracking with 2-d and 3-d models. Tech. rep., Digital Equipment Corporation, Cambridge Research Lab, 1997.

[11] Rohr, K. Human movement analysis based on explicit motion models. In *Motion-Based Recognition*. Kluwer Academic Publishers, Dordrecht Boston, 1997, ch. 8, 171–198.

[12] Sidenbladh, H. Black, M., and D.J., F. Stochastic tracking of 3d human figures using 2d image motion. In *Proc. of the 6th European Conference on Computer Vision, Dublin, 2000* (2000).

[13] Sullivan, J., Blake, A., Isard, M., and MacCormick, J. Object localization by bayesian correlation. In *Proc. 7th Int. Conf. on Computer Vision* (1999), vol. 2, 1068–1075.

[14] Vicon web based literature. URL http://www.metrics.co.uk, 1999. valid December 1999.

ISBN 0-7695-1272-0/01 $10.00 (C) 2001 IEEE