A Robust Multiple Hypothesis Approach to Monocular Human Motion Tracking

Cristian Sminchisescu, Bill Triggs

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Robust Multiple Hypothesis Approach to Monocular Human Motion Tracking

Cristian Sminchisescu, Bill Triggs

N° 4208

June 2001

THÈME 3

ISSN 0249-6399 ISRN INRIA/RR--4208--FR+ENG

A Robust Multiple Hypothesis Approach to Monocular Human Motion Tracking

Cristian Sminchisescu*, Bill Triggs †

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet MOVI

Rapport de recherche n° 4208 — June 2001 — 18 pages

Abstract: We study the problem of articulated 3D human motion tracking in monocular video sequences. Addressing problems related to unconstrained scene structure, uncertainty, and the high-dimensional parameter spaces required for human modeling, we present a novel, layered-robust, multiple hypothesis algorithm for estimating the distribution of the model parameters and propagating it over time. We use cost function based on robust contour and image intensity descriptors in a multiple assignment data association scheme. Our mixed discrete/global and continuous/local search technique uses both informed sampling and continuous optimization. Its novel hypothesis generation and pruning strategy focuses attention on poorly constrained directions in which large parameter space deviations are most likely, thus adaptively tracking the complex cost surface produced by non-linear kinematics, perspective projection and data-association problems. We also address the issue of semi-automatic acquisition of initial model pose and proportions, and show experimental tracking results involving complex motions with significant background clutter and self-occlusion.

Key-words: human motion analysis, particle filtering, high-dimensional search, constrained optimization

* INRIA Rhône-Alpes and INPG, *e-mail: Cristian.Sminchisescu@inria.fr*
† INRIA Rhône-Alpes and CNRS, *e-mail: Bill.Triggs@inria.fr*

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Une Approche Robuste Multi-Hypothèses Pour le Suivi de Mouvement Humain dans des Séquences Monoculaires

Résumé : On étudie le problème de suivi des mouvements articulaires 3D humain dans des séquences video monoculaires. En adresant des problèmes associés avec la structure non-contrainte de la scène, l'incertitude et l'espace paramétrique de grand dimension nécessaire pour la modélisation humaine, on présente un nouvel algorithme, robuste à l'échelle, pour l'estimation de la distribution sur les paramètres du modèle et leur propagation en temps. On utilise une fonction de côut basée sur des descripteurs de contour et d'intensité robustes, dans une stratégie d'appariement multiple. Notre approche de recherche hybride discrète/globale et continue/locale ultilise à la fois un échantionnage informé et une optimisation continue. Sa nouvelle statégie de génération et de sélection d'hypothèses, concentre ses efforts dans les directions mal estimées de l'espace paramétrique ou de grandes déviations dans les paramètres sont les plus probables, ainsi il sera possible de suivre adaptivement la surface de coût complexe produit par la cinématique non-linéaire, la projection perspective et le problème d'association de données dans l'image. On adresse aussi le problème d'acquisition semi-automatique des positions et proportions d'un modèle initiale et on présente des résultats expérimentaux avec des mouvements complexes, avec des fonds fortement perturbés et des auto-occultations.

Mots-clés : suivi de mouvement humain, optimisation contrainte, filtrage de particules, appariement robuste

Contents

1 Introduction

Extracting 3D human motion in realistic unconstrained environments based on monocular video sequences is a difficult process owing to the large number of parameters that even minimal generative human models necessarily have, the incomplete observability of monocular projection, and the ambiguities inherent in complex scene structure. We advocate an optimization based approach to this problem, the essence of which is hybrid local/global minimization of a cost function representing the posterior probability of the model given the data. Our cost function embodies the *generalized model-image mapping* that results from chaining the sequence of complex transformations that predict and evaluate model configurations in the image. It includes the effects of non-linear kinematics, perspective projection and model to image data association. In this context, the critical issues include strongly non-linear problem structure, observability and singularity issues, unconstrained motions, and the ambiguous assignment of model predictions to image observations, particularly in cluttered scenes and with inherently imperfect body and clothing modeling.

The main contribution of this paper is a robust, computationally efficient algorithm for estimating the posterior distribution over the model parameters and propagating it through time. As in other sampling-based approaches, the distribution is represented by a set of hypotheses together with a hypothesis generation and focusing mechanism. Our novel focusing mechanism captures the uncertainty inherent in the generalized model-image mapping. Our hybrid cost optimization algorithm involves both continuous and discrete components, namely: (i) local continuous optimization based on robust error distributions and incorporating joint limits; (ii) informed discrete sampling and hypothesis evaluation concentrated on the high-uncertainty parameter combinations of the continuous estimate; and (iii) a final hypothesis pruning and propagation stage. We also address the problem of semi-automatic model initialization from a single image, present a hierarchical algorithm for estimating body pose and proportions, and show experimental tracking results on video sequences involving complex motions with significant background clutter and self-occlusion.

1.1 Relation to Previous Work

Approaches to body and body part tracking using 3D models can be broadly classified as either deterministic continuous [7, 23, 18, 29, 28, 14], discrete [12, 19], or stochastic based on particle filtering [10, 9, 25] (see [13] for a comprehensive review). Continuous approaches typically use locally linearized model approximations and unimodal Gaussian error distributions, often propagated in time using extended Kalman filtering. Stochastic approaches use sample-based representations for the underlying distributions, priors based on various expected motion models, and search focusing methods ranging from importance sampling to annealing in an attempt to recover the 'typical set' of probable configurations.

As far as we know, no continuous approach has used either continuous optimization based on robust likelihood models, or optimizers incorporating hard joint limits (although [28] enforces joint limits at the level of the Kalman filter). Nor has any dealt with robustly extracted image descriptors, nor used a multiple assignment scheme, nor modeled or searched in a robust way the expected multimodal distribution over parameters.

Only a few works have addressed *monocular, 3D* tracking in a robust setting, particularly for full human body models. Deutscher [9] uses a MCMC technique based on annealing to sample the high dimensional space, but uses a clean background, an undressed person and silhouette and edge-based components to regularize his cost function. He tracks a walking person using annealing and temporal models but requires 3 cameras to track less periodic motion. Sidenbladh [25] uses a similar particle-based technique to track a walking person in a more complex setting, using an importance sampling method based on a strong learned prior motion model and an intensity-based cost function. Both of these approaches make strong assumptions, either on the simplicity of feature assignment (near-perfect silhouette and edge localization) and/or on the type of scene motions, to derive specialized imporance sampling functions.

Although the use of temporal models is appealing, any tracker designed to handle an unconstrained environment necessarily has to deal with irregular behavior, at best requiring several motion models and managing transitions between them. Unfortunately, multiple-model estimation schemes in high-dimensional spaces are at present very expensive computationally [5].

While not underestimating the importance of incorporating prior knowledge of the problem domain, in the present work we focus on techniques rooted in the generative model approach. We isolate two essential components of any optimization scheme based on a generative model, be it deterministic or stochastic. The first is the careful design of the cost function, which must robustly integrate the extracted image cues (contours, image intensity, motion boundaries) to limit the number of spurious local minima in parameter space. The second is the analysis of the generalized mapping linking model parameters to observations, which we believe is indispensable to allow accurate focusing of the search effort on the parameter space regions where good cost minima are most likely to occur. We argue that analysis of the uncertainty structure of this mapping is an effective generic technique for computationally tractable search in high-dimensional problems.

The paper is organized as follows: §2 briefly reviews the parameterization of our generative model; §3 presents the design of the cost function in terms of robust error distributions and image descriptors; §4 describes our hybrid continuous/discrete optimization method; §5 presents our semi-automatic initialization method; §6 discuses experimental tracking results; and §7 summarizes and discusses future research directions.

2 Generative Model

Our articulated model consists of kinematic and volumetric components. The underlying skeleton of the body is modeled in terms of kinematic chains associated body parts, represented using minimal parameterizations (x_a). The volumetric structure of the parts is modeled using superquadric ellipsoids, with additional tapering and bending parameters (x_d) for approximate surface modeling [1]. Each ellipsoid is discretized as a mesh in its topological domain Ω (see Figure 1).

A typical model has around 30 kinematic d.o.f. Supplementary parameters include internal proportions (x_i) (8 parameters encoding the positions of hip, clavicle or skull tip joints and 9 parameters for each volumetric part). Together, these produce a rather large parameter space, but only the kinematic parameters are actually estimated during tracking, although some of the internal proportions and volumetric parameters are estimated during model initialization (see Section 5). We believe that

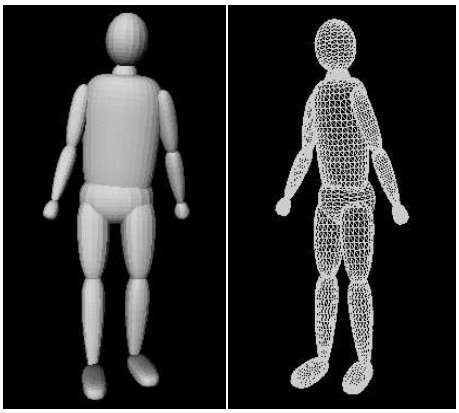

Figure 1: Human body model used during tracking

such a model offers several advantages in terms of high-level interpretation and occlusion prediction, and a good trade-off between complexity and coverage.

We encode the model representation in a single parameter vector $x = (x_a, x_d, x_i)$. Any node $u_i \in \Omega$ corresponding to one of the model parts, can be transformed into a 3D point $p_i = p_i(x)$, and subsequently into an image prediction $r_i = r_i(x)$ by means of a composite non-linear transformation:

$$r_i = T(x) = P(p_i = A(x_a, x_i, D(x_d, u_i))) \tag{1}$$

where D represents a sequence of parametric deformations which construct the corresponding part in a self-centered reference frame, A represents a chain of rigid transformations on the corresponding part kinematic chain, and P represents the perspective projection of the camera.

The process of model estimation involves a data association problem between individual model feature predictions r_i and one or more observations, that we shall generically denote $\bar{r}_i$ (with additional subscripts if these are several). We refer to $\Delta r_i(x) = \bar{r}_i - r_i(x)$ as the feature prediction error.

3 Cost Function Design

Whether continuous or discrete, any optimization process is dependent on the cost function to be minimized. Besides smoothness properties, we believe that cost functions should be designed to limit the number of spurious local minima in parameter space. Our approach employs a combination of edge and intensity information on top of a multiple assignment strategy based on a weighting scheme that focuses attention towards motion discontinuities. We also aim for a probabilistically interpretable model and build our cost function around robust error distributions

3.1 Error Distribution and Functional Form

Robust parameter estimates are intrinsically related to the choice of a realistic likelihood model that embodies the expected total inlier plus outlier distribution for the observation. We model the distribution in terms of robust radial terms, ρ_i, where $\rho_i(s)$ can be any increasing function with $\rho_i(0) = 0$ and $\frac{d}{ds}\rho_i(0) = \frac{\nu}{\sigma^2}$, that models an error distribution corresponding to a central peak, with influence σ, and a widely spread background of outliers ν. In this work, we use the following two robust error distributions (usually known as 'Lorentzian' and 'Leclerc' [4]):

$$\rho_i(s, \sigma) = \nu \log(1 + \frac{s}{\sigma^2}) \tag{2}$$

$$\rho_i(s, \sigma) = \nu(1 - e^{-\frac{s}{\sigma^2}}) \tag{3}$$

We aim towards a probabilistic interpretation and optimal estimates of the model parameters by maximizing the total probability according to Bayes rule:

$$p(x|\bar{r}) = \frac{1}{Z}p(\bar{r}|x)p(x) = \frac{1}{Z}\exp(-\int e(\bar{r}_i|x)di)p(x) \tag{4}$$

where $e(\bar{r}_i|x)$ is the cost density associated with the observation i, $p(x)$ is a prior on model parameters and Z a normalization constant. The cost for the observation i, expressed in terms of corresponding model prediction is $e(\bar{r}_i|x) = \frac{1}{N\nu}p_{ui}(x)$, where N is the total number of model nodes, W_i is a positive definite weighting matrix associated to the assignment i, and:

$$p_{ui}(x) = \begin{cases} \frac{1}{2}\rho_i(\Delta r_i(x) W_i \Delta r_i(x)^\top), & \text{if } i \text{ is assigned} \\ \nu_{bf} = \nu, & \text{if back-facing} \\ \nu_{occ} = k\nu, & k < 1, \text{ if occluded} \end{cases} \tag{5}$$

In our MAP approach, we discretize the continuous problem and attempt to minimize the negative log-likelihood for the total posterior probability, expressed as the following cost function:

$$\begin{aligned} f(x) &= -\log(p(\bar{r}|x)p(x)) \\ &= \frac{1}{N\nu}(\frac{1}{2}\sum_a \rho_a(\Delta r_a(x) W_a \Delta r_a(x)^\top) \\ &\quad + N_{bf}\nu_{bf} + N_{occ}\nu_{occ}) + f_p(x) \end{aligned} \tag{6}$$

where $f_p(x)$ is the negative log-likelihood of the prior, while N_{occ} and N_{bf} the occluded and backfaced (self-occluded) number of model nodes, respectively.

3.2 Image Descriptors

We choose both edge and intensity features for our cost function design. The images are smoothed with a Gaussian kernel, then a Sobel edge detector is applied. We employ a robust multiscale optical flow computation [3] to obtain both a flow field and an associated outlier map. The outlier map conveys useful information about the motion boundaries and is used to weight the significance of

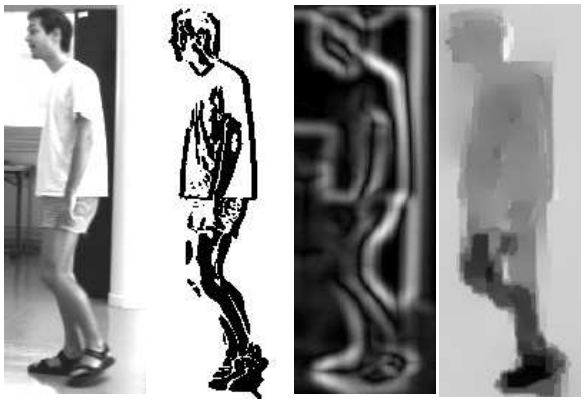

Figure 2: Image Processing Operations: original image, motion boundaries, edge detection, flow field in horizontal direction.

edges (see Figure 2). The motion boundaries are processed similarly to obtain a smooth image. For each node on model occluding contour, we perform a line search along the normal and retain all possible assignments within the search window, weighting them based on *both* the edge strength *and* their importance qualified by the motion boundary map. For nodes lying inside the object, we use intensity information derived from the robust optical flow. Including multiple edge assignments and flow, the first term in the cost function (6) becomes:

$$\begin{aligned} f(x) \quad = \quad & \tfrac{1}{2}(\textstyle\sum_i \sum_{a_e \in A} \rho_{i_{ae}}(\Delta r_{i_{ae}}(x) W_{i_{ae}} \Delta r_{i_{ae}}(x)^\top) \\ & + \textstyle\sum_j \rho_{j_{af}}(\Delta r_{j_{af}}(x) W_{j_{af}} \Delta r_{j_{af}}(x)^\top)) \end{aligned} \tag{7}$$

When evaluating sampled hypothesis, only the edge based term in (7) is used (see section 4.2).

4 Hybrid Optimization

Our search technique uses a combination of robust, consistent, local continuous optimization and more global (discrete) informed sampling.

4.1 Robust Bound Consistent Continuous Optimization

The model prediction in the image involves a complex generalized transformation including non-linear kinematics and perspective projection. The Jacobian matrix of this transformation, $J = \frac{\partial T}{\partial x}$ encodes the connection between differential quantities in the parameter and observation space, respectively.

The robust gradient and Hessian corresponding to the predicted model feature i and the assignments $a \in A$ can be derived:

$$g_i = J_i^\top \sum_{a \in A} \rho'_{ia} W_{ia} \Delta r_{ia} \tag{8}$$

$$H_i \approx J_i^\top \sum_{a \in A} (\rho_{ia}' W_{ia} + 2\rho_i''(W_{ia}\Delta r_{ia})(W_{ia}\Delta r_{ia})^\top) J_i \tag{9}$$

The gradient and Hessian corresponding to all observations are assembled, together with prior contributions:

$$g = g_o + \bigtriangledown f_p, \quad H = H_o + \bigtriangledown^2 f_p \tag{10}$$

For optimization, we use a second order damped Newton trust region method, where a descent direction is chosen by solving the regularized system [11]:

$$(H + \lambda W)\delta x = -g \tag{11}$$

where W is a symmetric positive-definite matrix and λ is a dynamically chosen weighting factor.

Joint constraints are handled in the optimizer, by projecting the gradient onto the current active constraint set. We find that imposing anatomical constraints on the joints effectively regularizes the tracking, providing an efficient way of bringing a weak form of prior knowledge into the problem. Also, notice that adding bound constraints in effect changes the shape of the cost function, and hence the minimum reached. In Figure 3 we plot a 1D slice through the constrained cost function together with a Taylor expansion based on the quadratic approximation. Notice nonzero gradient at minimum due to the presence of the bounds. The gradient changes abruptly because the active-set projection changes the motion direction during the slice to maintain consistency with the constraints.

4.2 Multiple Hypothesis Algorithm

We represent the evolution of our distribution over time as a set of hypotheses, weighted based on their costs. Although representational schemes based on propagating multiple hypotheses (samples) tend to increase the robustness in the estimated model, the great difficulty with high-dimensional distributions is hitting their typical sets (areas where most of the probability mass is concentrated). Consequently, informed ways of driving the sampling are needed. The main techniques developed in the vision sampling community involve either (i) a combination of careful regularization of the cost function, use of prior motion models, and importance sampling, or (ii) the adoption of a hierarchical search, actually a form of Gibbs sampling (assuming Gaussian marginals) referred as partitioned sampling [21], or (iii) general MCMC sampling techniques based on annealing.

Although we acknowledge the importance of both careful cost function design and the incorporation of more specific prior knowledge on human motion as a component of a robust system, we observe that no previous sampling technique has attempted to base its search on the fundamental, and therefore general *intrinsic* component of any generative model, the generalized transform

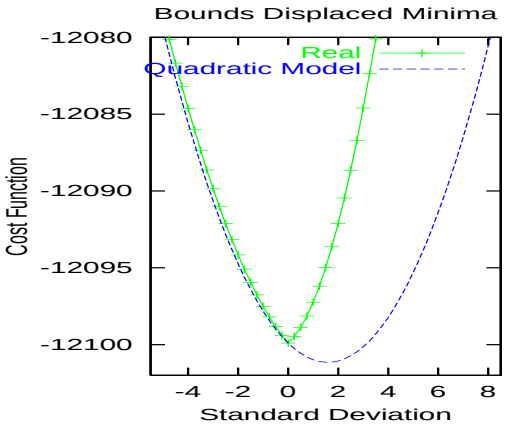

Figure 3: Displaced minimum due to bounds constraints

that maps points in model parameter space to observation-associated predictions, thus automatically including non-linear kinematics, perspective projection and data-association uncertainties and singularities. For example, such an algorithm is able to detect problems emphasized by [10] and to focus the search towards those directions in parameter space.

We propose a general algorithm in which each hypothesis over model configurations is subjected at each time-step to a robust continuous optimization. The estimation uncertainty of each hypothesis is represented by its covariance matrix (inverse Hessian) at its convergence point. An eigenvalue analysis primarily recovers the principal directions and we sample primarily along directions with high uncertainty, by normalizing according to their standard deviations. We generate and evaluate a set of hypotheses which is added to the hypothesis pool of the next frame. For speed, the evaluation uses only the edge component of the cost function, although some information from optical flow is automatically included by weighting by the confidence map derived from the motion boundaries. When all hypotheses have been processed, the pool is pruned by choosing the best k hypotheses which are then propagated to the next time step. We have empirically studied the form of the cost function by searching along uncertain directions for various model configurations. The careful selection of image descriptors, combined with the robust continuous optimization usually produces a smooth, singly peaked, distribution over model parameters. Locally multi-modal behavior is found in certain configurations as shown in Figure 5, which corresponds to one of the hypotheses at times 0.8sec and 0.9sec in the human tracking sequence of Figure (8). We have also studied the cost function along uncertain directions at much larger scales in parameter space (see Figure 6) and notice that we recover the expected robust shape of the distribution, without too many spurious local minima. Consequently the conjunction of robust cost function design and informed search is likely to be computationally efficient. The multiple hypothesis algorithm is briefly summarized in Figure (4).

```
k   number of hypothesis to propagate
n_s number of smallest singular vectors to consider
p_f perturbation factor along singular directions


MultipleHypothesisAlgorithm(k, n_s, p_f)
{
  t = 1
  hypothesisSet(t) = InitializeModel()
  for each time-frame (t)
  {
     for all h_c ε hypothesisSet(t)
     {
        nextSet = {}
        (h_i,H_i) = RobustContinousOptimize(h_c)
        (v, λ)_j=1..N = GeneralizedEigenDecomposition(H_i)

        for j = N downto N - n_s
        {
          σ_j = 1 / sqrt(λ_j)
          for p = -pf to p_f
          {
              nh_i = h_i + pσ_j v_j
              EvaluateCost(nh_i)
              Add(nh_i, nextSet)
          }
        }
     }
     hypothesisSet(t+1) = Prune(k, nextSet)

     t=t+1
  }
}
```

Figure 4: Multiple Hypothesis Algorithm

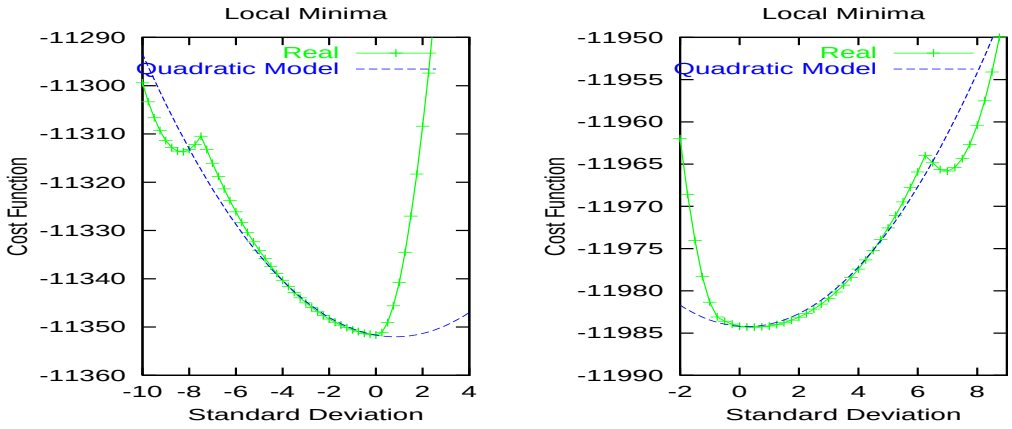

Figure 5: Local minima along the direction corresponding to the highest uncertainty, human sequence 0.8s and 0.9s

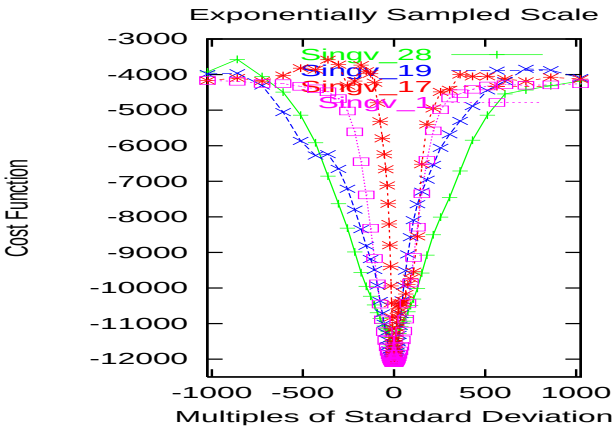

Figure 6: Cost function at larger scales in parameter space

5 Model Initialization

The visual tracking starts with a set of initial hypotheses resulting from a model initialization process. Correspondences need to be specified between model joint locations and approximate joint positions of the subject in the initial image. Given this input, we perform a consistent estimation of joint angles and body dimensions.

Previous approaches to model initialization from similar input and a single view, haven't fully addressed the generality and consistency problems [26, 2], not enforcing the joint limits constraints and making assumptions regarding either restricted camera models or restricted human subject poses in the image, respectively. Alternative approaches based on approximate pose recovery based on learned silhouette appearance [24], might be used to bootstrap an algorithm like the one we propose.

Our initialization algorithm is a hierarchical three step process that follows a certain parameter estimation scheduling policy. Each stage of the estimation process is essentially based on the formulation described in Section 4.1.

Hard joint limits constraints are automatically enforced at all stages by the optimization procedure, and corresponding parameters on the left and right sides of the body are "mirrored", while we collect measurements from the entire body (see below). Initialization proceeds as follows. Firstly, we solve an estimation problem under the given 3D model to 2D image correspondences (by minimizing the projection error), and prior intervals on the internal proportions and sizes of the body parts (namely parameters x_a, x_i and and some of x_d). Secondly we optimize only over the remaining volumetric body sizes alone (limb crossections and their tapering parameters x_d) while holding the other parameters fixed, using both the given correspondences and the local contour signal from image edges. Finally, we refine all model parameters (x) based on similar image information as in the second stage. The covariance matrix corresponding to the final estimate is used to generate an initial set of hypotheses, which are propagated in time using the algorithm described in (4.2). While the process is still heuristic, it gives a balance between stability and flexibility. In practice we find that enforcing the joint constraints, mirror information and prior bounds on the variation of body parameters gives far more stable and satisfactory results. However, in the monocular case, the initialization always remains uncertain in some directions, a problem that is probably only soluble by fusing pose information from multiple images.

6 Experiments

The experiments we show consists of an 8sec arm tracking sequence and a 1.2sec full body one in noisy image sequences with complex backgrounds and cluttered subjects as well as significant self-occlusion. Both sequences were shot at 25 FPS interlaced video rate. The experiments were run on a SGI O2 at 270 Mhz using an unoptimized implementation. They take around 10 sec/frame for the hand experiment and around 4 min/frame for the full human body sequence, most of the time being spent evaluating the cost function.

Shots showing the model (rendered) overlayed on the image are shown. For the arm sequence, only half is shown while the other half involves a similar reversed trajectory back to the starting position. The hand sequence involves 7 d.o.f which are tracked using just 3 hypothesis resulting from sampling along the smallest uncertain direction, while the full human body sequence involves 30 d.o.f which are tracked using 7 hypothesis based on discretizing the last 3 uncertain directions, each with 2 hypotheses, followed by pruning and propagation in each frame. The motion is successfully tracked over the entire sequence in both cases.

Notice that we are tracking in a cluttered background, with specular lighting and loose fitting clothing. As the hand sequence was shot with a camera situated quite close to the subject, the

deformations undergone by the arm muscles are significant. The imperfections in our arm model are also apparent. This caused problems when using a tracker based on a single hypothesis and non-robust Gaussian error distributions, especially at the points where the arm edges coincide with the pillar and when the arm self-occludes. Our focus of attention strategy based on motion boundaries helps to disambiguating the difficult configurations in which the arm (and the full human body), passes the white pillar with strong contrast edges. As emphasized in Section 4.2, we find multi-modal behavior during difficult configurations when the motions are far from parallel with the image plane, for instance between 2.2sec-4sec in the arm sequence, and over almost the entire full body sequence.

7 Conclusions

We have presented a novel multiple hypothesis approach to monocular human motion tracking. The approach is robust at various levels, namely: (1) in the cost function involving the extracted edge and intensity descriptors, and data association using a weighted assignment scheme based on focus of attention; (2) at the optimization level involving joint limit consistent continuous optimization based on realistic likelihood models, and at a more global scale by representing the posterior distribution as a discrete set of hypotheses. An important component of the method is an efficient hypothesis generation scheme, which automatically focuses the search in the areas of the parameter space that are poorly estimated due to the combined effect of non-linear kinematics, perspective projection and data-association non-linearities. We have also introduced a semi-automatic algorithm for pose and body proportion acquisition from a single image and reported experimental results in complex scenes. Future work will be directed towards a more rigorous analysis of the uncertainty during initialization and its propagation in the tracking stage, and to the inclusion and application of general priors on human poses and motions.

Acknowledgments

This work was supported by an EIFELL doctoral grant and the European Union under FET-Open project VIBES. We would like to thank Alexandru Telea for stimulating discussions and implementation assistance and Frédérick Martin for helping with the video capture and posing as a model.

References

[1] A.Barr, “Global and local deformations of solid primitives,” *Computer Graphics*, 18:21-30, 1984.

[2] C.Barron and I.Kakadiaris, “Estimating Anthropometry and Pose from a Single Image,” *CVPR*, pp. 669-676, 2000.

[3] M.Black and P.Anandan, “The Robust Estimation of Multiple Motions:Parametric and Piecewise Smooth Flow Fields,” *CVIU*, Vol.63, No.1, pp.75-104, 1996.

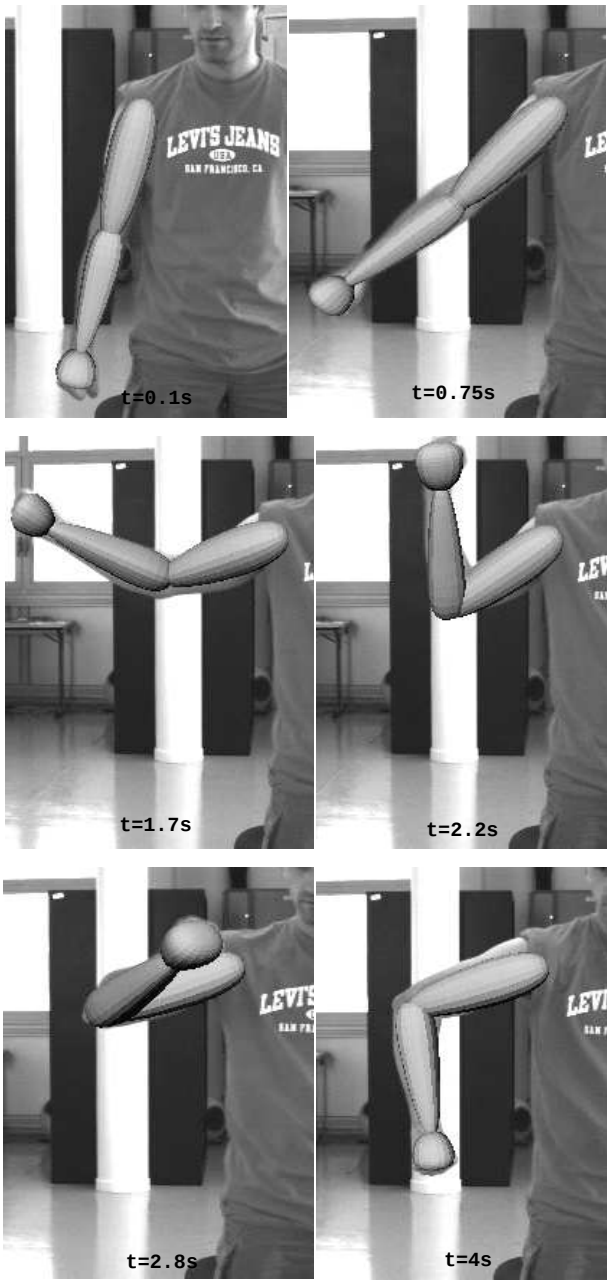

Figure 7: Shots in the arm tracking sequence

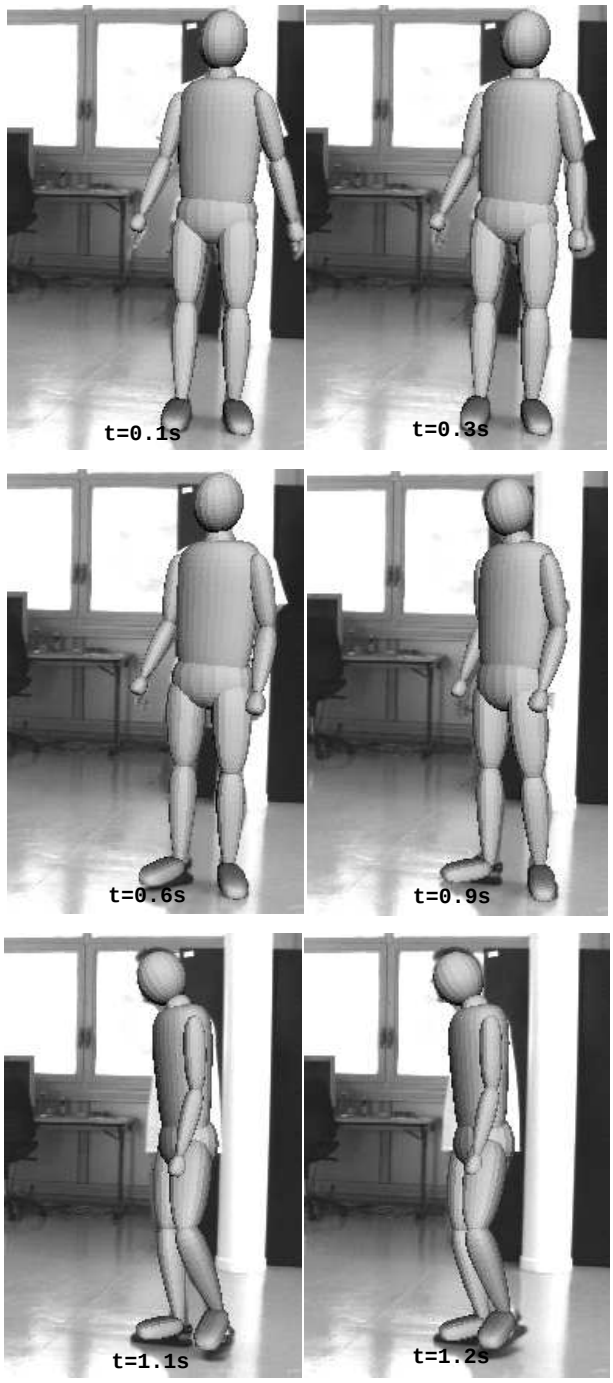

Figure 8: Shots during human body tracking sequence

Figure 9: Images in the human body tracking sequence

[4] M.Black, A.Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *IJCV*, Vol. 19, No. 1, pp. 57-92, July, 1996.

[5] A.Blake, B.North and M.Isard, "Learning multi-class dynamics", *ANIPS 11*, pp.389-395, 1999.

[6] M.Brand, "Shadow Puppetry," *ICCV*, pp.1237-1244, 1999.

[7] C.Bregler and J.Malik, "Tracking People with Twists and Exponential Maps," *CVPR*, 1998.

[8] T.Cham and J.Rehg, "A Multiple Hypothesis Approach to Figure Tracking", *CVPR*, Vol.2, pp.239-245, 1999.

[9] J.Deutscher, B.North, B.Bascle, A.Blake, "Tracking through Singularities and Discontinuities by Random Sampling," *ICCV*, pp.1144-1149, 1999.

[10] J.Deutscher, A.Blake, I.Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," *CVPR*, 2000.

[11] R.Fletcher, "Practical Methods of Optimization," *John Wiley*, 1987.

[12] D.Gavrila and L.Davis, "3-D Model Based Tracking of Humans in Action:a Multiview Approach," *CVPR*, pp. 73-80, 1996.

[13] D.Gavrila, "The Visual Analysis of Human Movement:A Survey," *CVIU*, Vol.73, No.1, pp.82-98, 1999.

[14] L.Gonglaves, E.Bernardo, E.Ursella, P.Perona, "Monocular Tracking of the human arm in 3D", *ICCV*, pp.764-770, 1995.

[15] T.Heap, D.Hogg, “Wormholes in Shape Space: Tracking through discontinuities changes in shape”, *ICCV*, pp.334-349, 1998.

[16] N.Howe, M.Leventon, W.Freeman, “Bayesian Reconstruction of 3D Human Motion from Single-Camera Video”, *ANIPS*, 1999.

[17] N.Jojic, M.Turk, T.Huang, “Tracking Self-Occluding Articulated Objects in Dense Disparity Maps,” *ICCV*, pp.123-130, 1999.

[18] I.Kakadiaris and D.Metaxas, “Model-Based Estimation of 3D Human Motion with Occlusion Prediction Based on Active Multi-Viewpoint Selection,” *CVPR*, pp. 81-87, 1996.

[19] J.Kuch and T.Huang “Vision-based modeling and tracking for virtual teleconferencing and telecollaboration,” *ICCV*, pp.666-671, 1995.

[20] M.Leventon and W.Freeman, “Bayesian Estimation of 3-d Human Motion from an Image Sequence,”, *TR-98-06*, MERL, 1998.

[21] J.MacCormick and M.Isard, “Partitioned sampling, articulated objects, and interface-quality hand tracker”, *ECCV*, Vol.2, pp.3-19, 2000.

[22] D.Morris and J.Rehg, “Singularity Analysis for Articulated Object Tracking”, , pp. 289-296, CVPR 1998.

[23] J.Rehg and T.Kanade “Model-Based Tracking of Self Occluding Articulated Objects,” *ICCV*, pp.612-617, 1995.

[24] R.Rosales and S.Sclaroff, “Inferring Body Pose without Tracking Body Parts,” *CVPR*, pp.721-727, 2000.

[25] H.Sidenbladh, M.Black, D.Fleet, “Stochastic Tracking of 3D Human Figures Using 2D Image Motion,” *ECCV*, 2000.

[26] C.J.Taylor, “Reconstruction of Articulated Objects from Point Correspondences in a Single Uncalibrated Image,” *CVPR*, pp.677-684, 2000.

[27] B.Triggs, P.McLauchlan, R.Hartley, A.Fitzgibbon, “Bundle Adjustment - A Modern Synthesis,” *Vision Algorithms: Theory and Practice*, Springer-Verlag, LNCS 1883, 2000.

[28] S.Wachter and H.Nagel, “Tracking Persons in Monocular Image Sequences,” *CVIU*, 74(3):174-192, 1999.

[29] C.Wren and A.Pentland, ”DYNAMAN; A Recursive Model of Human Motion,” *MIT Media Lab Tech. Report*, No. 451.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifi que
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399