Human Pose Estimation from Silhouettes. A Consistent Approach Using Distance Level Sets

Cristian Sminchisescu, Alexandru Telea

HUMAN POSE ESTIMATION FROM SILHOUETTES A CONSISTENT APPROACH USING DISTANCE LEVEL SETS

C. Sminchisescu[1] **and A. Telea**[2]

[1] INRIA-Rhone-Alpes
655 avenue de l'Europe, 38330 Montbonnot, France
Cristian.Sminchisescu@inria.fr

[2] Eindhoven University of Technology
Department of Mathematics and Computer Science
Den Dolech 2, 5600 MB Eindhoven, The Netherlands
alext@win.tue.nl

ABSTRACT

We present a novel similarity measure (likelihood) for estimating three-dimensional human pose from image silhouettes in model-based vision applications. One of the challenges in such approaches is the construction of a model-to-image likelihood that truly reflects the good configurations of the problem. This is hard, commonly due to the violation of consistency principle resulting in the introduction of spurious, unrelated peaks/minima that make the search for model localization difficult. We introduce an entirely continuous formulation which enforces model estimation consistency by means of an attraction/explanation silhouette-based term pair. We subsequently show how the proposed method provides significant consolidation and improved attraction zone around the desired likelihood configurations and elimination of some of the spurious ones. Finally, we present a skeleton-based smoothing method for the image silhouettes that stabilizes and accelerates the search process.

Keywords: human tracking, model-based estimation, constrained optimization, level set methods, fast marching methods

1 INTRODUCTION AND PREVIOUS WORK

Human pose estimation from images is an active area of computer vision research with many potential applications ranging from computer interfaces to motion capture for character animation, biometrics or intelligent surveillance. One promising approach, called *model based* [Smin01b, Deut00, Heap98, Smin01a, Gavr96, Breg98, Kakad96, Rehg95], relies on a 3D articulated volumetric model of the human body to constrain the localization process in one or several images. The goal in human pose estimation applications is to estimate the model's articulation and possibly structural parameters such that the projection of the 3D geometrical model closely fits a human in one or several images. Typically, model localization is a multi-dimensional expensive search process in the model parameter space for good cost configurations defined in terms of maxima of a *likelihood*, or minima of an energy function. Such costs are defined in terms of the association of model predictions with extracted image features. The search process produces a parameter configuration which brings the 3D model close to the tracked 2D image in the metric of the predefined likelihood model. The above problem is hard since likelihood cost surfaces are typically multi-peaked, due to factors like multiple scence objects, ambiguous feature assignments, occlusions, and depth uncertainties.

Search strategies for locating good peaks in the model parameter space based on local and global search methods, possibly in temporal sequences, have received significant attention [Smin01b, Deut00, Heap98, Smin01a, Gavr96, Breg98] and are not addressed here. However, the dificulty and intinsically ill-posed nature of such search problems raise two complementary questions about the design of the cost surface whoose minima are to be found:

- what are *good image features* which will readily qualify for likelihood terms for sampling and continuous evaluation ?

- how to *define* such terms to limit the number of spurious minima in parameter space and render the search more efficient and effective.

Likelihood models defined in terms of edges [Deut00, Smin01a, Kakad96], silhouettes [Deut00, Smin01a, Heap98] or intensities [Smin01a, Side00, Rehg95] are the most common. While image intensities seem to be good cues for various types of optical-flow based local search, they are not invariant to lighting changes, and typically rely on low inter-frame intensity variations and motion. It is consequently difficult to sample configurations out of the region where such photometric model is valid. Edges and/or silhouettes have therefore been more used in approaches that employ, at least partially, some form of parameter-space sampling [Deut00, Smin01a, Heap98].

Deutscher [Deut00] uses a silhouette based term for his cost function design in a multi-camera setting. However, this term peaks if the model is inside the silhouette without demanding that the silhouette area is fully explained (see Sec. 4.1). Consequently, an entire family of undesired configurations situated inside the silhouette will generate good costs under this likelihood model. Moreover, the term is purely discrete, not suitable for continuous estimation. The situation is alleviated by the use of the additional cues and sensor-fusion from multiple cameras with good results. Delamarre [Dela99] uses silhouette contours in a multi-camera setting and computes assignments using a form of ICP (Iterative Closest Point) algorithm and knowledge of normal contour directions. The method is local and not necessarily enforces globally consistent assignments, but again relies on fusing information from many camera to ensure consistency. Brand [Bran99] and Rosales [Rosa00] use silhouettes to infer temporal and static human poses. However, their motivation is slightly different in using silhouettes as inputs to a system which directly learns 3D to 2D mappings.

Summarizing, many likelihood terms used in model-based vision applications have the undesirable property that they not only peak around the desired model configurations, which correspond to subject localization in the image, but also in totally unrelated, false configurations. This poses huge burdens on any search algorithm, as the number of spurious minima could grow unbounded and therefore discriminating them from "good peaks" can only be done via temporal processing. Consequently, any finite samples/hypothesis estimator has a great chance to miss significant, true minima.

In practice, extracting pose from silhouette using single images remains an under-constrained problem with potential multiple solutions. A more global search method, multiple cameras, temporal disambiguation and/or additional features have thus to be used in conjunction with the local method we propose in this work, to robustify the search for good cost configurations [Smin01b, Deut00, Heap98, Smin01a, Gavr96]. In this paper, we assume a reasonable initialisation and restrict our attention to the design of likelihoods with larger basin of attraction zones and globally consistent responses around the desirable cost minima. We achieve this by means of an entirely continuous formulation and a new likelihood term for silhouettes in model-based applications. The proposed term allows a globally consistent response for the subject localization in the image by means of a pair of attraction/explanation components that a) push the geometric model inside the subject's silhouette and b) demand that the area associated with the silhouette is entirely explained by the model. We subsequently show how this proposal significantly improves the pose estimation results compared to previously used similarity measures.

In Section 2, we describe the human body model we employ. Section 3 outlines the search process for optimal configurations. Section 4 introduces our new likelihood terms and details its two components. Section 5 presents a new technique for smoothing the image-acquired silhouettes that stabilizes and accelerates the search process. Finally, Section 6 concludes the paper and proposes directions for future work.

2 HUMAN MODEL

2.1 MODEL DESCRIPTION

Our human body model (Fig.1) consists of kinematic 'skeletons' of articulated joints controlled by angular **joint parameters** x_a, covered by 'flesh' built from superquadric ellipsoids with additional tapering and bending parameters [Barr84]. A typical model has around 30 joint parameters, plus 8 **internal proportion** parameters x_i encoding the positions of the hip, clavicle and skull tip joints, plus 9 **deformable shape** parameters for each body part, gathered into a vector x_d. The state of a complete model is thus given as a single parameter vector $x = (x_a, x_d, x_i)$. We note, however, that only joint parameters are typically estimated during object localization and tracking, the other parameters remaining fixed.

Although this model is far from photo-realistic, it suffices for a high-level interpretation and realistic occlusion prediction. Moreover, it offers a good trade-off between computational complexity and coverage

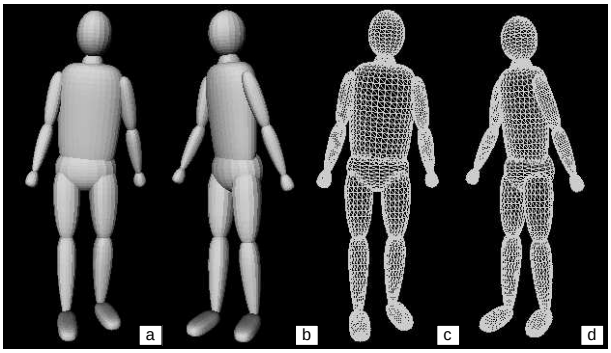

Figure 1: Human model: flat shaded (a,b) and discretization (c,d)

in typical motion tracking applications.

2.2 MODEL TO IMAGE FITTING

The model is used in the human pose estimation application as follows (see also Fig. 2 for an overview of the application pipeline).

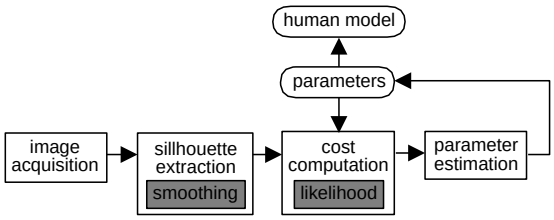

Figure 2: Human pose estimation application pipeline

The pipeline starts by extracting a human silhouette (see the example in Fig. 3 b) from the camera-acquired images (Fig. 3 a) by subtracting the scene background and thresholding the result to a bilevel image. To stabilize the further parameter estimation step, a special smoothing is applied on the extracted image. This smoothing is described separately in Sec. 5. The model's superquadric surfaces are discretized as meshes parameterized by angular coordinates in a 2D topological domain. Mesh nodes u_i are transformed into 3D points $p_i = p_i(x)$ and then into predicted image points $r_i = r_i(x)$ using composite nonlinear transformations $r_i(x) = T_i(x) = P(p_i(x)) = P(A(x_a, x_i, D_p(x_d, u_i)))$, where D_p is a sequence of parametric deformations that construct the corresponding part in its own reference frame, A is a chain of rigid transformations that map it through the kinematic chain to its 3D position, and P is the perspective projection.

During parameter estimation (see Sec. 3), prediction-to-image matching cost metrics are evaluated for predicted image feature r_i, and the results are summed to produce the image contribution to the overall parameter space cost function. For certain likelihood terms like edge based ones, predictions r_i are associated with nearby image features $\bar{r}_i$. The cost is then a function of the prediction errors $\Delta r_i(x) = \bar{r}_i - r_i(x)$. For other likelihood terms (like the silhouette attraction term we employ here), a potential surface is built for the current image, and the prediction is only evaluated at a certain location on this surface.

3 PARAMETER ESTIMATION

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

$$p(x|\bar{r}) \propto p(\bar{r}|x)p(x) = \exp\{-(e_a + e_s)\}p(x) \quad (1)$$

where e_a and e_s are the new silhouette likelihood terms we propose, defining similiarity criteria between the model projection and the image silhouette to be defined in the next section, and $p(x)$ is a prior on model parameters. The prior encodes static knowledge on humans, such as anatomical joint angle limits for parameters or non-penetration constraints on the body parts (see [Smin01b, Smin01a] for details).

In a maximum a-posteriori estimate (MAP) approach, we spatially discretize the continuous formulation in Eqn. 1, and attempt to minimize the negative log-likelihood, or 'energy', for the total posterior probability. The energy is expressed as the following cost function:

$$\begin{aligned} f(x) &= -\log(p(\bar{r}|x)p(x)) = -\log p(\bar{r}|x) - \log p(x) \\ &= e_a + e_s + f_p(x) \end{aligned}$$

where $f_p(x)$ is the negative log of the model prior. In the following, we shall concentrate on the behavior and properties of the negative log-likelihood $e_a + e_s$.

Various search methods attempt to identify the minima of the function f, by either local continous descent, stochastic search, parameter space subdivision or combinations of them [Smin01b, Deut00, Heap98, Smin01a, Gavr96, Breg98]. All these methods require the evaluation of f. Continuous methods require supplementary evaluations of the first order gradient g and sometimes the second order Hessian H of f. In this paper, we use a second order local continuous method, where a descent direction is chosen by solving the regularized subproblem [Flet87]:

$$(H + \lambda W)\delta x = -g, \text{ subject to } C_{jl} \cdot x < 0$$

where:

- W is a symmetric positive-definite stabilization matrix (ofter set to identity)

- λ is a dynamically chosen weighting factor
- C_{jl} is a matrix containing joint angle limits constraints acting as effective priors, defining an addmissible subspace to search for model parameters (see [Smin01b, Smin01a] for details).

The parameter λ controls the descent type: $\lambda \to \infty$ leads to a gradient descent, while $\lambda \to 0$ leads to a Newton-Raphson step. The optimization routine automatically decides over the type and size of the optimal step within the admissible trust radius (see [Flet87, Trig00] for details).

4 OBSERVATION LIKELIHOOD

Whether continuous or discrete, the search process depends critically on the observation likelihood component of the parameter-space cost function. Besides smoothness properties, necessary for the stability of the local continuous descent search, the likelihood should be designed to limit the number of spurious local minima in parameter space. We propose a new likelihood term, based on two components:

- the first component maximizes the model-image silhouette area overlap.
- the second component pushes the model inside the image silhouette.

The above pair of cost terms produces a global and consistent response. In other words, this term enforces the model to remain within the image silhouette, but also demands that the image silhouette is entirely explained, i.e. that *all* silhouette parts contribute to the cost function that drives the fitting process. In the following, we detail the two cost components.

4.1 SILHOUETTE-MODEL AREA OVERLAP TERM

This term maximizes the model-image area overlap. The area of the predicted model can be computed from the model's projected triangulation by summing over all visible triangles $t \in V_t$ (triangles having all the vertices $(x_i, y_i)_{i=1..3}$ visible).

$$S_a = \sum_{t \in V_t} \sum_{i=1}^{3} (x_{i \odot 3}(y_{i+1 \odot 3} - y_{i+2 \odot 3}) \qquad (2)$$

where $\odot$ describes the modulo operation, and the computation assumes the triangle vertices are sorted in counter-clockwise order to preserve positive area sign. In subsequent derivations we drop the modulo notation for simplicity.

Let S_g be the area of the target silhouette. The area alignment cost, i.e. the difference between the model and image silhouette areas, is:

$$e_a = \frac{1}{2\sigma^2}(\sum_{t \in V_t} S_a - S_g)^2 \qquad (3)$$

The gradient and Hessian for the area-based cost-term can subsequently be derived (by dropping the scaling term):

$$g_a = \frac{de_a}{dx} = (\sum_{t \in V_t} S_a - S_g) \sum_{t \in V_t} \frac{\partial S_a}{\partial x} \qquad (4)$$

where:

$$\frac{\partial S_a}{\partial x} = \sum_{i=1}^{3} \frac{\partial x_i}{\partial x}^{\top} (y_{i+1} - y_{i+2}) \qquad (5)$$

$$+ \sum_{i=1}^{3} x_i (\frac{\partial y_{i+1}}{\partial x} - \frac{\partial y_{i+2}}{\partial x})^{\top} \qquad (6)$$

$$H_a = \frac{d^2 e_a}{dx^2} = \sum_{t \in V_t} \frac{\partial S_a}{\partial x}^{\top} \frac{\partial S_a}{\partial x} \qquad (7)$$

$$+ (\sum_{t \in V_t} S_a - S_i) \sum_{t \in V_t} \frac{\partial^2 S_a}{\partial x^2} \qquad (8)$$

One should notice that the individual partial derivatives $\frac{\partial x_i}{\partial x}$ and $\frac{\partial y_i}{\partial x}$ represent the columns of the individual Jacobian matrix evaluated at the corresponding prediction for the mesh node i, $r_i(x) = (x_i, y_i)$. In practice, computing node visibility and area differences is rather fast, as we use the frame and z buffers to this end.

4.2 SILHOUETTE ATTRACTION TERM

This second term pushes the model inside the image silhouette. Adding over all projected model nodes i, this term writes:

$$e_s = \frac{1}{2\sigma^2} \sum_i e_{s_i} \qquad (9)$$

where e_{s_i} is the distance from a predicted model point $r_i(x)$ to a given silhouette S_g. We estimate e_{s_i} by computing the distance transform D of the silhouette S_g and evaluating it in the points i:

$$e_{s_i}(r_i(x), S_g) = D(r_i(x)) \qquad (10)$$

We use a level-set based approach to quickly and robustly estimate D, as follows. We initialize D to zero on S_g, i.e. regard S_g as the zero level set of the function D. Next, we compute D by solving the Eikonal equation [Seth99]:

$$|\nabla D| = 1 \tag{11}$$

for all points outside S_g. The solution of equation 11 has the property that its isolines, or level sets, are at equal distance from each other in the 2D space (Fig. 3). Consequently, D is a good approximation of the distance transform D.

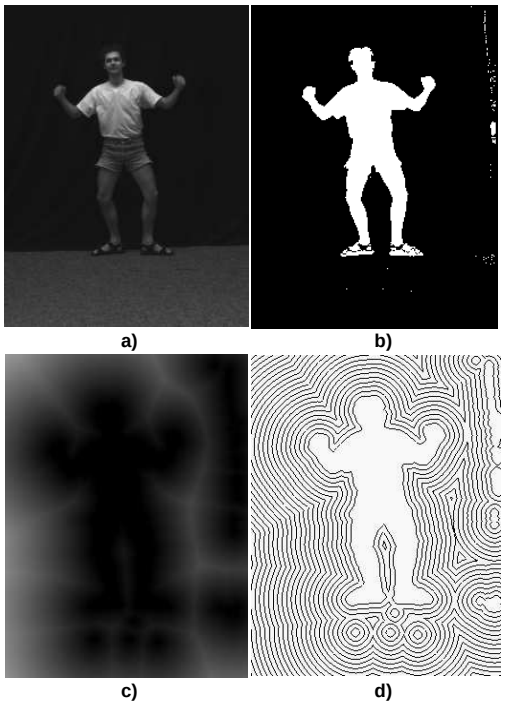

Figure 3: Distance transform computation: original image (a), silhouette (b), distance plot (c) and distance level sets (d)

Equation 11 is efficiently solved by using the fast marching method (FMM), introduced by Sethian in [Seth96]. We briefly outline here the FMM. A detailed description of the FMM, up to the implementation details we have ourselves used, is given in [Seth96, Seth99]. First, D is initialized to zero in all points on the silhouette S_g. Next, the solution D is built outwards starting from the smallest known D value. This is done by evolving a so-called *narrow band* of pixels, initially identical to S_g, in normal direction to S_g, with unit constant speed. As the narrow band evolves, it takes the shape of the consecutive, equidistant level sets, or isolines, of the function D (Fig. 3 d).

Using the FMM to compute the distance D has several advantages. First, the function D obtained is continuous over the 2D plane, which is important as we need to evaluate its first and second order derivatives, as explained below. Secondly, the FMM performs robustly even for noisy silhouettes S_g. This is essential for practical applications, as the silhouettes extracted from real images have many disconnected, spurious pixels (Fig. 3 b is a typical example). Thirdly, the FMM is very efficient, as it needs $O(n * logk)$ operations, where n is the number of image pixels and k is the average number of pixels in the narrowband, of the same order as the number of pixels on the silhouette's contour. D is computed in real time for 500^2 pixel images on an SGI O2 R5000 machine. Finally, implementing the FMM is straightforward, as described in [Seth96]. Overall, we believe that using the FMM to compute D is a more efficient and effective method than e.g. chamfer based methods widely used in vision and imaging applications.

The gradient and Hessian of the corresponding silhouette attraction term are computed from the model-image Jacobian, as follows:

$$g_s = \sum_i \frac{dD(r_i(x))}{dx} = \sum_{i \in V} J_i^\top \frac{\partial D}{\partial r_i} \tag{12}$$

$$H_s = \sum_i \frac{d^2 D}{dx^2} \approx \sum_{i \in V} J_i^\top \frac{\partial^2 D}{\partial r_i^2} J_i \tag{13}$$

Figure 4 shows the effect of the silhouette attraction and area overlap terms for two images taken from a longer tracking sequence. The figure shows the initial images (a,e), the initial model configuration (b,f), and the fitting results obtained when using only the silhouette attraction term (c,g) and finally both the silhouette attraction and the area overlap terms (d,h). One can notice that the silhouette attraction term does not suffice for a good fit. Indeed, any parameter configuration which places the model inside the image silhouette can be potentially chosen. Adding the area overlap term stabilizes the estimation and drives it towards relatively satisfactory results. Moreover, the cost term has the desired properties of a wide attraction zone. This makes it a good candidate in tracking applications where recovery from tracking failures is highly desirable.

5 SILHOUETTE SMOOTHING

The gradient g and Hessian H introduced in the previous sections are at the core of the optimization process that fits the model to the observed image features. The stability of the optimization is influenced by the behavior of g and H: if the silhouette data are noisy, then the cost terms e_s and e_a, and their derivatives

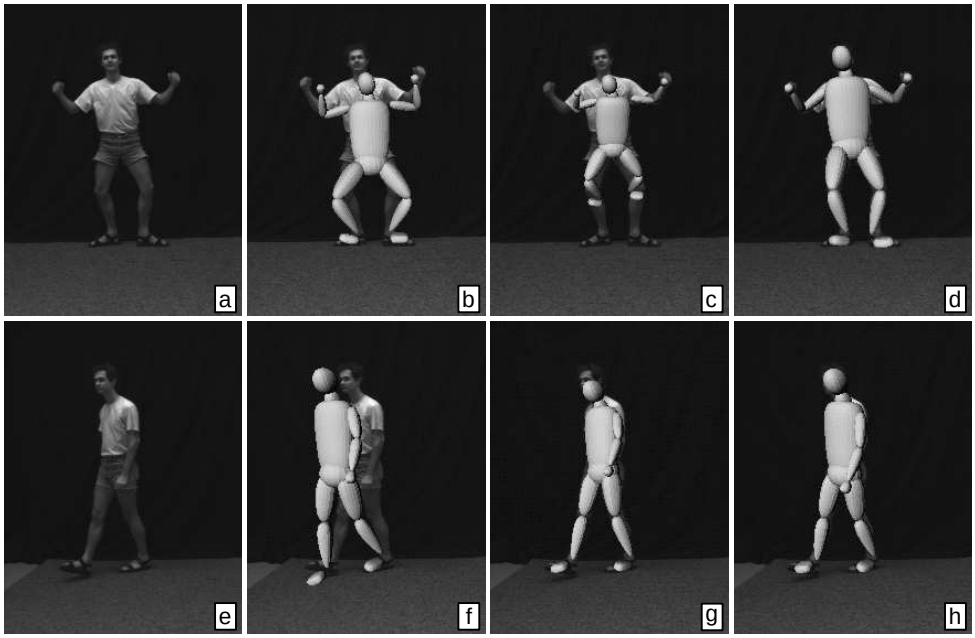

Figure 4: Model estimation based on various silhouette terms original images (a,e), initial models (b,f), silhouette attraction term only (c,g), silhouette attraction and area overlap terms (d,h)

g and H, are not smooth functions. In such cases, the optimization process might fail or take too long to converge or might fit the model erroneously to the image silhouette.

We alleviate this problem by performing a *smoothing* on the silhouettes acquired from the image data. The smoothing aims to produce silhouettes from the image data that can be easier approximated by our human body models than the original 'raw' silhouettes. The process runs as follows (see Fig. 6 for an overview).

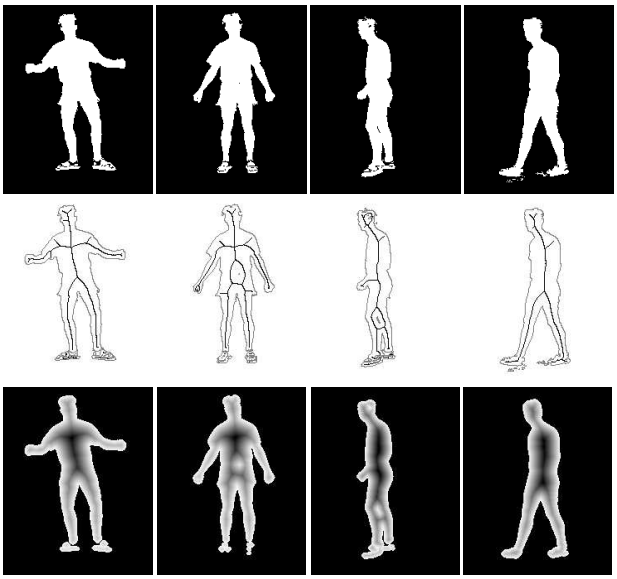

Figure 5: Examples of raw silhouettes, skeletons, and smoothed silhouettes

First, the raw silhouettes are extracted from the image data, as explained in 4. Due to the limitations of the extraction process, these silhouettes may have a jagged boundary, contain spurious pixels, or miss pixels on the real silhouette, as in Fig. 6 b.

In second next step, we compute the *skeleton* of the silhouette, as follows. We apply the FMM algorithm inwards on the raw silhouette and compute the distance map D_1 of all the points inside the silhouette to its boundary (Fig. 6 b). The silhouette skeleton is then computed as being those points of the evolving narrow band that meet other similar points due to the band's evolution under normal speed. In other words, the skeleton points are those points where the narrow band collapses onto itself during its evolution driven by the FMM algorithm. We identify these points using a technique similar to the ones described in [Sidd99, Ogni95b].

In the third step, the obtained skeleton (Fig. 6 c) is pruned of its small, less significant branches by retaining only its points that originate from points on the initial narrow band situated at a distance larger than a given threshold [Ogni95a, Ogni95b, Sidd99]. The above pruning scheme is based on two observations: a) every skeleton point is generated by the collapsing of a compact segment of the original boundary [Sidd99, Kimm95], and b) the importance of a skeleton point can be measured by the length of the boundary segment out of which it originates [Ogni95a, Ogni95b].

In the last step, we 'inflate' the pruned skeleton to ob-

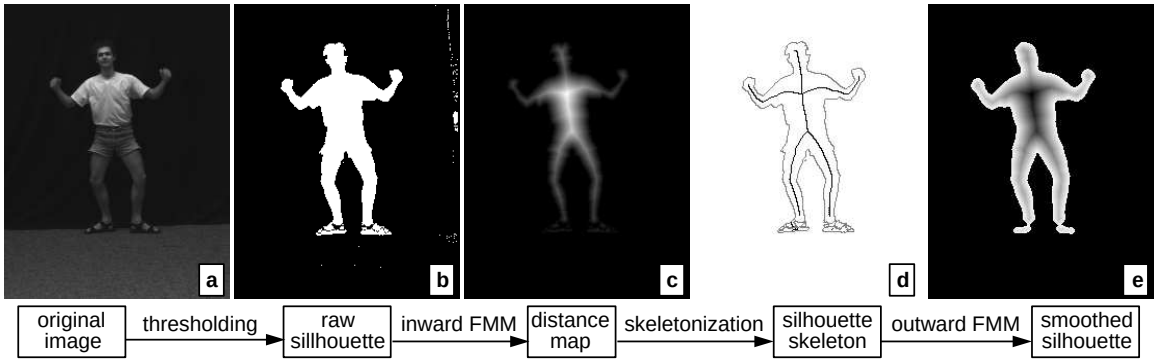

Figure 6: Skeleton-based silhouette smoothing pipeline

tain the smoothed silhouette. To do this, we execute again the FMM algorithm outwards from the skeleton, as follows. We initialize the narrow band to the skeleton points and the function D to the value of $-D_1$ at those points, where D_1 is the distance from the skeleton to the silhouette, computed in the previous step. We stop the FMM execution when points of the outwards evolving narrow band reach a D value of zero. At that moment, the inflated skeleton matches the initial silhouette (Fig. 6 d). However, due to its pruning, most of the noise of the initial raw silhouette has been removed, as seen in the examples in Fig. 5.

Since the FMM algorithm performs in real time, as noted in Sec. 4.2, the whole skeleton-based smoothing process takes less than a second for our typical images. By adjusting the skeleton pruning threshold, we obtain different smoothing levels. Smoother silhouettes, produced by a higher threshold, lead in practice to a more stable and sensibly faster convergence of the model parameter estimation. Moreover, pruned skeletons typically lead, due to the properties of the Eikonal equation used in the reconstruction, to silhouettes having rounded edges. These shapes are easier appproximated by the superquadric shapes used in our human body model than the raw, arbitrarily shaped silhouettes. However, if the skeletons are pruned too much, the smoothed silhouettes might miss important image cues, such as the orientation of a limb. Conversely, less smoothed silhouettes are closer to the observed data, thus more accurate, but, as mentioned, may lead to numerically unstable derivative estimations. Currently we estimate, by trial and error, a good value for the pruning threshold for a given application configuration (camera parameters, lighting, raw silhouette extraction parameters, optimization method parameters, etc). This value works well for the various images we have tried it on. However, a better strategy we plan to investigate is to use an adaptively optimal threshold for each image.

6 CONCLUSIONS

We have presented a method to build more consistent likelihood terms for silhouettes, and applied it for human pose estimation in a model based context. Aiming to build cost surfaces whose minima accurately reflect the good configurations in the problem, we define a novel likelihood model composed of an attraction term and an area overlap term which ensures consistent model localization in the image with improved attraction zones. Secondly, we propose a smoothing method for the silhouettes extracted from the image data that stabilizes the optimization process used for pose estimation. Both the likelihood attraction term and silhouette smoothing method are based on distance functions extracted using level-set techniques for evolving boundaries under constant speed in the normal direction. In particular, the fast marching method allows us to calculate distance transforms, skeletons, and to reconstruct silhouettes from their skeletons in a simple to implement and efficient way.

Our future work aims at employing silhouette skeletons, extracted with level set methods, directly as likelihood terms for human pose estimation applications. Together with this, we aim to develop an automatic procedure of setting the prunung threshold for the skeleton-based smoothing we employ on the image-extracted silhouettes.

REFERENCES

[Barr84] Barr, A.: *Global and local deformations of solid primitives*, Computer Graphics, no.18, pp. 21–30, 1984.

[Bran99] Brand, M.: *Shadow Puppetry*, Proc. ICCV, pp.1237–1244, 1999.

[Breg98] Bregler, C. and Malik, J.: *Tracking People with Twists and Exponential Maps*, Proc. CVPR, 1998.

[Dela99] Delamarre, Q. and Faugeras, O.: *3D Articulated Models and Multi-View Tracking with Silhouettes*, Proc. ICCV, pp. 716–721, 1999.

[Deut00] Deutscher, J. and Blake, A. and Reid, I.: *Articulated Body Motion Capture by Annealed Particle Filtering*, Proc. CVPR, vol. 2, pp. 126–133, 2000.

[Flet87] Fletcher, R.: *Practical Methods of Optimization*, John Wiley & Sons, 1987.

[Gavr96] Gavrila, D. and Davis, L.: *3-D Model Based Tracking of Humans in Action:a Multi-view Approach*, Proc. CVPR, pp. 73–80, 1996.

[HAWG] Hanim-Humanoid Animation Working Group, *Specifications for a standard humanoid*, available at *http://www.h-anim.org/Specifications/H-Anim1.1/*

[Heap98] Heap, T. and Hogg, D.: *Wormholes in Shape Space: Tracking through discontinuities changes in shape*, Proc. ICCV, pp.334–349, 1998.

[Howe99] Howe, N. and Leventon, M. and Freeman, W.: *Bayesian Reconstruction of 3D Human Motion from Single-Camera Video*, Proc. ANIPS, 1999.

[Kakad96] Kakadiaris, I. and Metaxas, D.: *Model-Based Estimation of 3D Human Motion with Occlusion Prediction Based on Active Multi-Viewpoint Selection*, Proc. CVPR, pp. 81–87, 1996.

[Kimm95] Kimmel, R. and Shaked D. and Kiryati N. and Bruckstein A. M.: *Skeletonization vis Distance Maps and Level Sets*, Computer Vision and Image Understanding, vol. 62, no. 3, pp. 382–391, 1995.

[MacC00] MacCormick, J. and Isard, M.: *Partitioned sampling, articulated objects, and interface-quality hand tracker*, Proc. ECCV, vol.2, pp.3–19, 2000.

[Ogni95a] Ogniewicz, R. L.: *Automatic Medial Axis Pruning by Mapping Characteristics of Boundaries Evolving under the Euclidean Geometric Heat Flow onto Voronoi Skeletons*, Harvard Robotics Laboratory, Technical Report 95–4, 1995.

[Ogni95b] Ogniewicz, R. L. and Kubler. O: *Hierarchic Voronoi Skeletons*, Pattern Recognition, nr. 28, pp. 343–359, 1995.

[Rehg95] Rehg, J. and Kanade, T.: *Model-Based Tracking of Self Occluding Articulated Objects*, Proc. ICCV, pp.612–617, 1995.

[Rosa00] Rosales, R. and Sclaroff, S.: *Inferring Body Pose without Tracking Body Parts*, Proc. CVPR, pp.721–727, 2000.

[Seth96] Sethian, J. A.: *A Fast Marching Level Set Method for Monotonically Advancing Fronts*, Proc. Nat. Acad. Sci. vol. 93, nr. 4, pp. 1591–1595, 1996.

[Seth99] Sethian, J. A.: *Level Set Methods and Fast Marching Methods, Cambridge University Press*, 2nd edition, 1999.

[Sidd99] Siddiqi, K. and Bouix, S. and Tannenbaum, A. and Zucker, S. W.: *The Hamilton-Jacobi Skeleton*, Proc. Intl. Conf. on Computer Vision ICCV '99, pp. 828–834, 1999.

[Side00] Sidenbladh, H. and Black, M. and Fleet, D.: *Stochastic Tracking of 3D Human Figures Using 2D Image Motion*, Proc. ECCV, pp. 702–718, 2000.

[Smin01a] Sminchisescu, C. and Triggs, B.: *A Robust Multiple Hypothesis Approach to Monocular Human Motion Tracking*, research report INRIA-RR-4208, June 2001.

[Smin01b] Sminchisescu, C. and Triggs, B.: *Covariance-Scaled Sampling for Monocular 3D Body Tracking*, Proc. CVPR, pp. 447–454, 2001.

[Smin01c] Sminchisescu, C. and Telea, A.: *A Framework for Generic State Estimation in Computer Vision Applications*, Proc. ICVS, Springer Verlag, 2001.

[Sull99] Sullivan, J. and Blake, A. and Isard, M. and MacCormick, J.: *Object Localization by Bayesian Correlation*, Proc. ICCV, pp. 1068–1075, 1999.

[Terz88] Terzopoulos, D. and Witkin, A. and Kass, M.: *Constraints on deformable models: Recovering 3-D shape and non-rigid motion*, Artificial Intelligence, 36(1), pp. 91–123, 1988.

[Trig00] Triggs, B. and McLauchlan, P. and Hartley, R. and Fitzgibbon, A.: *Bundle Adjustment - A Modern Synthesis*, Vision Algorithms: Theory and Practice, Springer-Verlag, LNCS 1883, pp. 298–372, 2000.

[Wach99] Wachter, S. and Nagel, H.: *Tracking Persons in Monocular Image Sequences*, Proc. CVIU, 74(3), pp. 174–192, 1999.

[Wren] Wren, C. and Pentland, A.: *DYNAMAN; A Recursive Model of Human Motion*, MIT Media Lab technical report No. 451, 2000.

[Zhu97] Zhu, S. C. and Mumford, D.: *Learning Generic Prior Models for Visual Computation*, IEEE Trans. PAMI, 19(11), pp 1236–1250, 1997.