



**HAL**  
open science

## Consistency and Coupling in Human Model Likelihoods

Cristian Sminchisescu

► **To cite this version:**

Cristian Sminchisescu. Consistency and Coupling in Human Model Likelihoods. 5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '02), May 2002, Washington, United States. pp.27–32, 10.1109/AFGR.2002.1004125 . inria-00590164

**HAL Id: inria-00590164**

**<https://inria.hal.science/inria-00590164>**

Submitted on 3 May 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Consistency and Coupling in Human Model Likelihoods

Cristian Sminchisescu

INRIA Rhône-Alpes, 655 avenue de l'Europe, 38330 Montbonnot, France  
*Cristian.Sminchisescu@inrialpes.fr; www.inrialpes.fr/movi/people/Sminchisescu*

## Abstract

*This paper presents two novel likelihood terms for silhouettes and contours in model-based contexts. Despite the power of such formulations, building likelihoods that truly reflect the good configurations of the problem is by no means easy due to, most commonly, the violation of consistency principle resulting in the introduction of spurious, unrelated peaks/minima that make target localization difficult. We introduce an entirely continuous formulation which enforces consistency by means of an attraction/explanation pair for silhouettes. For contours, we address the search window vs. noise level dilemma by means of a combined robust estimation and feature coupling solution which builds a likelihood model not only in terms of individual contour responses but also "Gestalt" type higher-order couplings between matched configurations. We subsequently show how the proposed method provides significant consolidation and improved attraction zone around the true cost minima and elimination of some of the spurious ones.*

**Keywords:** *robust matching, Markov random fields, model-based estimation, human tracking, constrained optimization*

## 1 Introduction

This paper addresses the problem of building likelihood terms for problems involving object localization and tracking, in the model-based context, i.e. in situations where an approximate geometric model of the object is available. We pay particular attention to 3D human model-based estimation and likelihood models defined in terms of silhouette [5, 4] and edge [14, 5, 10] features, as they readily qualify for both sampling and continuous optimization. In human model-based estimation, one has to address two major difficulties:

(I) Search the model parameter space for good cost configurations over a likelihood surface. Such strategies have been proposed in the continuous, discrete and mixed cases [14, 5, 3, 9, 7, 2] and are not addressed in this work.

(II) Design of the likelihood surface in terms of implicit or explicit association of model predictions with extracted image features. It is known that such cost surfaces are typically multi-modal and in vision this owes to factors like feature assignment, occlusions and depth ambiguities.

Clearly, the two difficulties (I) and (II) interact. Over-simplified likelihoods defined inconsistently or based on de-

coupled local image features, undesirably respond not only around the model configurations corresponding to subject localization in the image, but also in a potentially un-bounded family of spurious (but good cost!) peaks. Additional undesirable side-effects include singular, flat or very low curved resulting minima. Such likelihood surfaces create huge problems to any search algorithm, as the number of spurious minima could grow unbounded and discriminating them from 'good peaks' can only be done via temporal processing. Nevertheless, in such a context, the chance of any finite samples/hypothesis estimator to miss significant, true minima increases substantially.

In this paper, we propose an *entirely continuous formulation* and two novel likelihood models for contours and silhouettes which directly address these problems. The *silhouette likelihood* term we derive allows a globally consistent response for the object localization in the image by means of a pair of attraction/explanation terms that not only attracts the model inside the subject silhouette, but also demands that the area associated with the silhouette is entirely explained by the model.

The *contour likelihood* term is designed to deal with clutter and the data association problems encountered in model-based contour identification. Our approach preserves the 3D top-down, continuous and independent nature of model-based search process, but augments the likelihood term, with non-local coupled 'Gestalt' response terms (besides the 'classical' individual contour responses), to account for the properties of the matched configurations (symmetry, smoothness), in the spirit of bottom-up MRF methods [8, 11, 12, 17].

We subsequently experimentally show that more convenient silhouette and contour likelihood models can be built in the proposed way, in the sense of allowing both wider attraction/search zones and elimination or down-weighting of the minima associated with spurious, un-consistent configurations.

### 1.1 Previous Work

**Silhouette Likelihoods** Deutscher [5] uses a discrete silhouette based term for his cost function design in a multi-camera setting. The term peaks if the model is inside the silhouette without demanding that the silhouette area is fully ex-

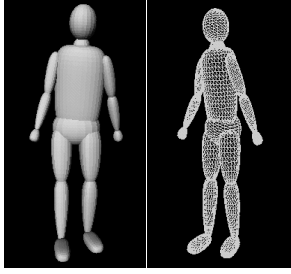


Figure 1: Human model (a) flat shaded (b) wireframe

plained (see Sec. 4.1.1). Consequently, an entire family of un-informative configurations situated inside the silhouette will generate good costs under this likelihood model. The situation is alleviated by the use of additional edge cues and multiple cameras. Delamarre [4] uses silhouette contours in a multi-camera setting and computes assignments using an Iterative Closest Point algorithm and knowledge of normal contour directions. The method is local and doesn't enforce globally consistent assignments, but relies on fusing information from many cameras to ensure consistency. Brand [1] and Rosales [13] 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.

**Contour Likelihoods** Generating, detecting or matching image contours for a particular object class is difficult. *Bottom-up MRF methods* [8, 11, 12, 17] define an energy functional that encodes the basic properties of the contour (smoothness, edge responses) and solve using dynamic programming, or Monte-Carlo methods for more general types of couplings. These methods are elegant, but lead to expensive high-dimensional feature space representations for which finding good proposal distributions is non-trivial. Their generalization to 3D and the invariance related issues seem also problematic. *Top-down model based approaches* [15, 5, 14], use a valid predicted model configuration (smoothness properties are built into the model parameterization) to generate independent local search processes in the neighborhood of individual model predictions. Matching configurations is fast, but the matching process is essentially local, so arbitrary wrong global results due to gradual loss in model regularization effect are obtained with the increase in the search window.

## 2 Generative Model

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. A typical model has around 30 joint parameters, plus 8 **inter-**

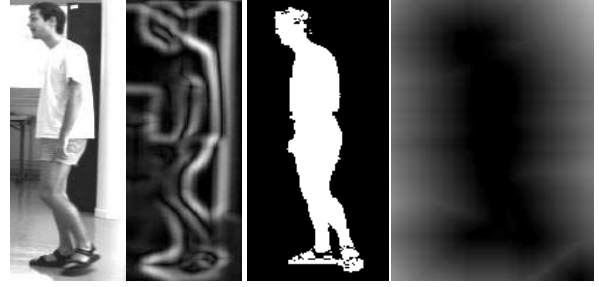


Figure 2: Image features: (a) original image (b) intensity-edge energy (c) silhouette (d) distance transform of corresponding silhouette

**nal proportion** parameters  $\mathbf{x}_i$  encoding the positions of the hip, clavicle and skull tip joints, plus 9 **deformable shape** parameters for each body part, gathered into a vector  $\mathbf{x}_d$ . A complete model can be encoded as a single large parameter vector  $\mathbf{x} = (\mathbf{x}_a, \mathbf{x}_d, \mathbf{x}_i)$ . During localization or tracking we typically estimate only joint parameters. The model is used as follows. 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  $\mathbf{r}_i = \mathbf{r}_i(\mathbf{x})$  using composite nonlinear transformations  $\mathbf{r}_i(x) = P(\mathbf{p}_i(\mathbf{x})) = P(A(\mathbf{x}_a, \mathbf{x}_i, D(\mathbf{x}_d, \mathbf{u}_i)))$ , where  $D$  represents a sequence of parametric deformations that construct the corresponding part in its own reference frame,  $A$  represents a chain of rigid transformations that map it through the kinematic chain to its 3D position, and  $P$  represents perspective image projection. During model estimation, prediction-to-image matching cost metrics are evaluated for predicted image feature  $\mathbf{r}_i$ , 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  $\mathbf{r}_i$  are associated with nearby image features  $\bar{\mathbf{r}}_i$ . The cost is then a (typically robust) function of the prediction errors  $\Delta\mathbf{r}_i(\mathbf{x}) = \bar{\mathbf{r}}_i - \mathbf{r}_i(\mathbf{x})$ . For other likelihood terms, a potential surface is built for the current image, and the prediction is only evaluated at a certain location on the potential surface.

## 3 Problem Formulation

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

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

where  $e_a, e_s, e_c$  are likelihood terms corresponding to silhouette and edge cues costs (potentials), to be defined in the next section,  $p(\mathbf{x})$  is a prior on model parameters. 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:

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

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

Any search method requires the evaluation of  $f$  while continuous methods require supplementary first order gradient  $\mathbf{g}$  and sometimes second order Hessian  $H$  operators. For optimization we use a second order trust region method, where a descent direction is chosen by solving the regularized sub-problem [6]:

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

where  $\mathbf{W}$  is a symmetric positive-definite matrix,  $\lambda$  is a dynamically chosen weighting factor and  $\mathbf{C}_{jl}$  is a matrix containing joint angle limits constraints (see [14] for details). Note that the multiple minima search problem is not addressed herein (see [14, 5, 3, 9, 7] for this) but all continuous and discrete evaluation tools (likelihood terms, their first and second order operators) are derived in this work.

## 4 Observation Likelihood

### 4.1 Silhouette Likelihood

This cost term is based on a pair of sub-components, one which pushes the model inside the image silhouette, while the other maximizes the model-image silhouette area overlap. The cost term is global and consistent, in that it is not only enforcing the model remains within the image silhouette, but also demands that the image silhouette is entirely explained.

#### 4.1.1 Silhouette-Model Area Overlap Term

The area of the predicted model can be computed from the model projected ‘‘triangulation’’ by summing over all ‘‘visible’’ triangles  $t \in V_t$  (triangles having all the vertices  $(x_i, y_i)_{i=1,2,3}$  visible).

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

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 writes:

$$e_a = \frac{1}{2\sigma^2} \left( \sum_{t \in V_t} S_t - S_g \right)^2$$

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

$$\mathbf{g}_a = \frac{de_a}{d\mathbf{x}} = \left( \sum_{t \in V_t} S_t - S_g \right) \sum_{t \in V_t} \frac{dS_t}{d\mathbf{x}} \quad (3)$$

where:

$$\frac{dS_t}{d\mathbf{x}} = \sum_{i=1}^3 \frac{dx_i}{d\mathbf{x}}^\top (y_{i+1} - y_{i+2}) + \sum_{i=1}^3 x_i \left( \frac{dy_{i+1}}{d\mathbf{x}} - \frac{dy_{i+2}}{d\mathbf{x}} \right)^\top$$

$$\mathbf{H}_a = \frac{d^2 e_a}{d\mathbf{x}^2} = \sum_{t \in V_t} \frac{dS_t}{d\mathbf{x}}^\top \frac{dS_t}{d\mathbf{x}} + \left( \sum_{t \in V_t} S_t - S_g \right) \sum_{t \in V_t} \frac{d^2 S_t}{d\mathbf{x}^2} \quad (4)$$

where:

$$\begin{aligned} \frac{d^2 S_t}{d\mathbf{x}^2} &= \sum_{i=1}^3 \frac{d^2 x_i}{d\mathbf{x}^2} (y_{i+1} - y_{i+2}) + \sum_{i=1}^3 x_i \left( \frac{d^2 y_{i+1}}{d\mathbf{x}^2} - \frac{d^2 y_{i+2}}{d\mathbf{x}^2} \right)^\top \\ &\quad + 2 \sum_{i=1}^3 \frac{dx_i}{d\mathbf{x}}^\top \left( \frac{dy_{i+1}}{d\mathbf{x}} - \frac{dy_{i+2}}{d\mathbf{x}} \right) \end{aligned}$$

One should notice that the individual partial derivatives  $\frac{dx_i}{d\mathbf{x}}$  and  $\frac{dy_i}{d\mathbf{x}}$  represent the columns of the individual Jacobian matrix evaluated at the corresponding prediction for the mesh node  $i$ ,  $\mathbf{r}_i(\mathbf{x}) = (x_i, y_i)$ . In practice computing node visibility and area differences is rather fast calculation which we perform using the z-buffer.

#### 4.1.2 Silhouette Attraction Term

This term pushes the model inside the image silhouette. Adding over all projected model nodes  $i$  the cost writes:

$$e_s = \frac{1}{2\sigma^2} \sum_i e_{s_i} \quad (5)$$

The distance from a predicted model point  $\mathbf{r}_i(\mathbf{x})$  to a given silhouette  $S_g$  can be written as a quadratic Chamfer distance, between model prediction  $\mathbf{r}_i(\mathbf{x})$  and points  $\mathbf{s}_i$  on the given silhouette  $S_g$ :

$$e_{s_i}(\mathbf{r}_i(\mathbf{x}), S_g) = \min_{\mathbf{s}_i \in S_g} \|\mathbf{r}_i(\mathbf{x}) - \mathbf{s}_i\| \quad (6)$$

It is known that such Hausdorff distance is giving a good quadratic approximation ([16]). The Chamfer distance can be computed fast by means of dynamic programming [7] (note

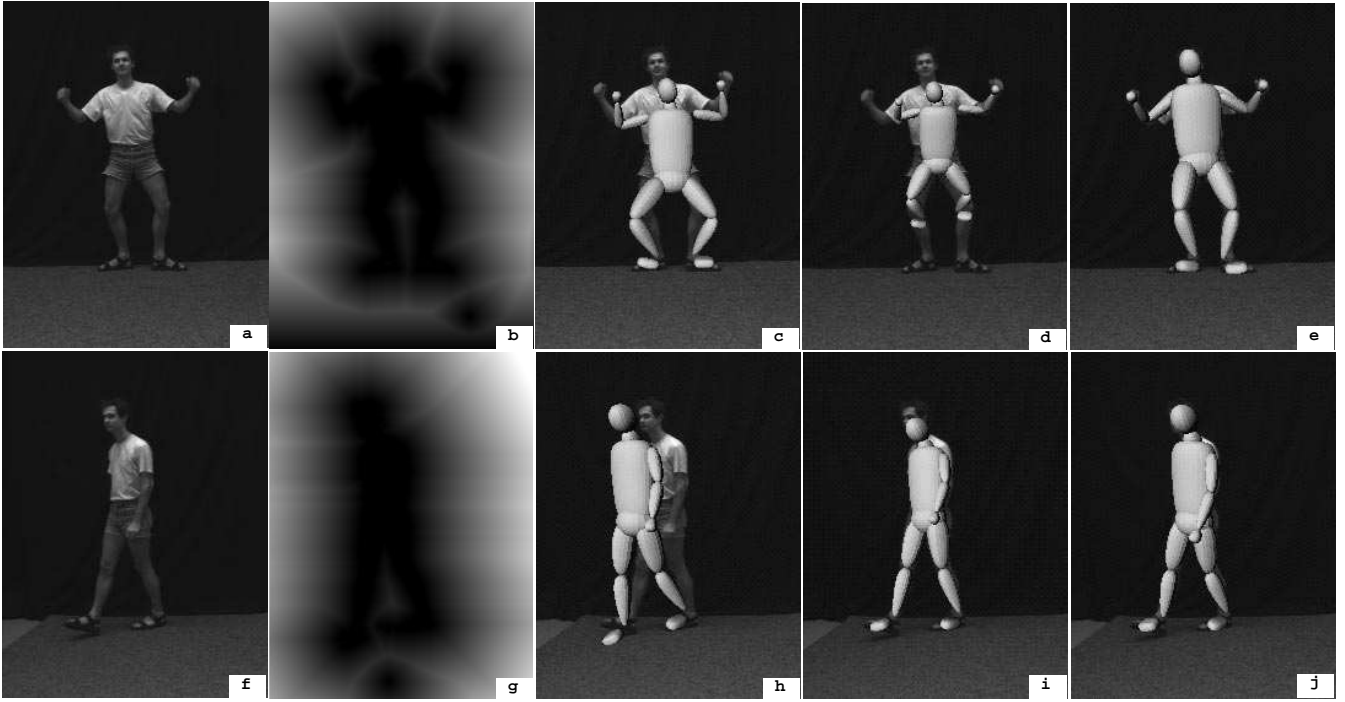


Figure 3: Model estimation based on various silhouette terms original images (a,f), their distance transforms (b,g), initial models (c,h), silhouette attraction term only (d,i), silhouette attraction and area overlap terms (e,j). Silhouette attraction term produces inconsistent likelihoods. The pair of attraction/explanation terms produces good fits and enforces consistency

that [7, 16] employed this distance for edges and in a discrete evaluation context). However, the distance lacks immediate continuous structure. Consequently, given a Chamfer silhouette image (see fig.3 b and fig.3 g), we build a 2-dimensional continuous potential surface  $P_s$  by fitting local quadric surfaces to  $3 \times 3$  image patches. The gradient and Hessian of the corresponding cost term can therefore be derived from the model-image Jacobian, and the corresponding quadric terms:

$$\mathbf{g}_s = \sum_i \frac{dP_s(\mathbf{r}_i(\mathbf{x}))}{d\mathbf{x}} = \sum_{i \in V} \mathbf{J}_i^\top \frac{dP_s}{d\mathbf{r}_i} \quad (7)$$

$$\mathbf{H}_s = \sum_i \frac{d^2 P_s}{d\mathbf{x}^2} \approx \sum_{i \in V} \mathbf{J}_i^\top \frac{d^2 P_s}{d\mathbf{r}_i^2} \mathbf{J}_i \quad (8)$$

**Experiments** In fig.3 we show, for two image silhouettes, the initial image (fig.3 a and fig.3 f), the distance transform of the corresponding silhouette (fig.3 b and fig.3 g), the initial model configuration (fig.3 c and fig.3 h) and the fitting results obtained when using only the silhouette attraction term (fig.3 d and fig.3 i) and finally both the silhouette and the area overlap term (fig.3 e and fig.3 j). One can notice that, as expected, the silhouette attraction terms does not suffice for a good fit and potentially any configuration which places the model inside the image silhouette can be chosen. On the contrary, the area overlap term stabilizes the estimation and

drives it towards rather satisfactory results. Also, note that the cost term also has the desired properties of a wide attraction zone, being thus a good candidate in tracking applications where recovery from tracking failure is a highly desirable property.

## 4.2 Contour Likelihood

For edge extraction, the images are smoothed with a Gaussian kernel, contrast normalized, and a Sobel edge detector is applied. For visible nodes on model occluding contours ( $\mathcal{O}$ ), we perform line search along the projected contour normal in order to match the corresponding prediction with an image edge. As the search and matching processing for each model prediction is performed independently, it is not guaranteed that the matched image features form admissible configurations. Robust estimation is one way this problem can be alleviated. There are however, two problems with it. The first is caused by nearby outliers whose influence can't be entirely switched off by the robust cost. This is particularly apparent in contexts where the inter-frame motion is sufficiently important that a larger noise process has to be used. Secondly, the independence assumptions during search define an incorrect (or actually weaker) model to robustify and robustness can't tackle this.

Consequently the idea we follow here is to use robust es-

timation but to provide it with the correct model of what the target to be localized is (reflecting the non-independent relations-couplings between model elements). We particularize this in the case of human motion analysis for the models of limbs, but the idea applies far more generally in the sense that the semantics between any model predicted points has to be preserved between individual corresponding image matched features.

#### 4.2.1 Robust Error Distributions

Robust parameter estimation can be viewed as the choice of a realistic total likelihood model for the combined inlier and outlier distributions for the observation. We model the total likelihood in terms of robust radial terms  $\rho_i$ , where  $\rho_i(s)$  can be any increasing function with  $\rho_i(0) = 0$  and  $\frac{d}{ds}\rho_i(0) = \frac{\nu}{\sigma^2}$ . These model error distributions corresponding to a central peak with scale  $\sigma$ , and a widely spread background of outliers  $\nu$ . Here we used the ‘Lorentzian’  $\rho_i(s, \sigma) = \nu \log(1 + \frac{s^2}{\sigma^2})$  and ‘Leclerc’  $\rho_i(s, \sigma) = \nu(1 - \exp(-\frac{s^2}{\sigma^2}))$  robust error potentials.

The cost of individual observation  $i$ , expressed in terms of corresponding model prediction is  $e(\bar{\mathbf{r}}_i|\mathbf{x}) = \frac{1}{N\nu}e_{ui}(\mathbf{x})$ , where  $N$  is the total number of model nodes,  $\mathbf{W}_i$  is a positive definite weighting matrix associated to the assignment  $i$ , and:

$$e_{ui}(\mathbf{x}) = \begin{cases} \frac{1}{2}\rho_i(\Delta\mathbf{r}_i(\mathbf{x})\mathbf{W}_i\Delta\mathbf{r}_i(\mathbf{x})^\top), & \text{if } i \text{ is assigned} \\ \nu_{bf} = \nu, & \text{if back-facing} \\ \nu_{occ} = k\nu, & k < 1, \text{ if occluded} \end{cases}$$

#### 4.2.2 Matching Consistency and Data Coupling

The matching consistency term exploits the property that a configuration of matched points has to have similar appearance with the model prediction of the corresponding points. While this property might appear trivially fulfilled for model-driven search, independence assumptions and wider noise models might generate wrong matched configurations with good cost. Properties like symmetry (of a limb, for instance) can be violated during the search process. The uniqueness assignment principle is also, no longer guaranteed.

We consequently propose building likelihood terms that encode not only the probabilities of edge responses but also the model symmetries or other non-independent model properties. For instance for any predicted model node  $\mathbf{r}_i(x)$ , lying on the model occluding contour, there exist a symmetric node (with respect to a projected limb axis)  $\mathbf{r}_{i_c}(x)$  in the model. One can derive a symmetry corrected negative log-likelihood term, based on the deviation between the model predicted symmetries and the matched ones:  $\Delta\mathbf{r}_{i_c} = (\mathbf{r}_i(\mathbf{x}) - \mathbf{r}_{i_c}(\mathbf{x})) - (\bar{\mathbf{r}}_i - \bar{\mathbf{r}}_{i_c})$ . The term is essentially exploiting the semantics present in the model which has to directly

transfer to the results of correspondence process:

$$e_c = \frac{1}{2}\left(\sum_{i \in \mathcal{O}} \rho_i(\Delta\mathbf{r}_i(\mathbf{x})\mathbf{W}_i\Delta\mathbf{r}_i(\mathbf{x})^\top) + \sum_{i \in \mathcal{O}} \varphi_{i_c}(\Delta\mathbf{r}_{i_c}(\mathbf{x})\mathbf{W}_{i_c}\Delta\mathbf{r}_{i_c}(\mathbf{x})^\top)\right)$$

where  $\varphi$  is a negative log-likelihood of a robust error potential with a wide attraction zone. The gradient and Hessian corresponding to the predicted assignments and couplings, can be derived using the columns of the model-image Jacobian matrix,  $\mathbf{J}_i = \frac{d\mathbf{r}_i}{d\mathbf{x}}$ :

$$\begin{aligned} \mathbf{g}_c &= \sum_{i \in \mathcal{O}} \mathbf{J}_i^\top \rho'_i \mathbf{W}_i \Delta\mathbf{r}_i + \sum_{i \in \mathcal{O}} \mathbf{J}_i^\top \varphi'_{i_c} \mathbf{W}_{i_c} \Delta\mathbf{r}_{i_c} \\ \mathbf{H}_c &\approx \sum_{i \in \mathcal{O}} \mathbf{J}_i^\top (\rho_i' \mathbf{W}_i + 2\rho_i''(\mathbf{W}_i \Delta\mathbf{r}_i)(\mathbf{W}_i \Delta\mathbf{r}_i)^\top) \mathbf{J}_i \\ &\quad + \sum_{i \in \mathcal{O}} \mathbf{J}_i^\top (\varphi_{i_c}' \mathbf{W}_{i_c} + 2\varphi_{i_c}''(\mathbf{W}_{i_c} \Delta\mathbf{r}_{i_c})(\mathbf{W}_{i_c} \Delta\mathbf{r}_{i_c})^\top) \mathbf{J}_i \end{aligned}$$

For a human body model, the above ‘‘symmetrized’’ edge term will add over pairs of symmetric visible nodes lying on occluding contours for all limbs (the increase in cost is not significant, for  $n$  edge based-terms an extra  $\mathcal{O}(n/2)$  terms are added). In fact other properties present in the model that could be potentially lost during correspondence search, could be built in order to drive the matching process towards valid configurations (like for instance further ‘‘Gestalt’’ properties like collinearity, cocircularity, etc.). The coupling of model and observation nodes can be of arbitrary order, although with the expected increase in computational cost. The problem can be formulated, as above, such that the continuous properties of the likelihood model are preserved. For instance collinearity properties for nodes on limbs can be built into a determinant and require  $\mathcal{O}(n^2)$  terms. Conic properties involve deriving  $2nd$  order parametric curves, etc.

**Experiments** We built symmetry constraints into the likelihood model. The first experiment we show involves passing a lower arm over an image arm. As one can see in fig.5, the cost has multiple minima, owing to incorrect assignments of both model edges to the same image edge. The problem is indeed alleviated by the introduction of modeling constraints on the matched configurations. A similar experiment is performed for the motion of a foot (fig.6), and the pure independent correspondence search gets stuck into a local minima owing to incorrect assignments due to the break in model symmetries.

## 5 Conclusions and Future Work

We have presented a method to build consistent likelihood models for contours and silhouettes, which can be applied when a model of the object to be tracked or localized is available, and studied several applications in the context of human

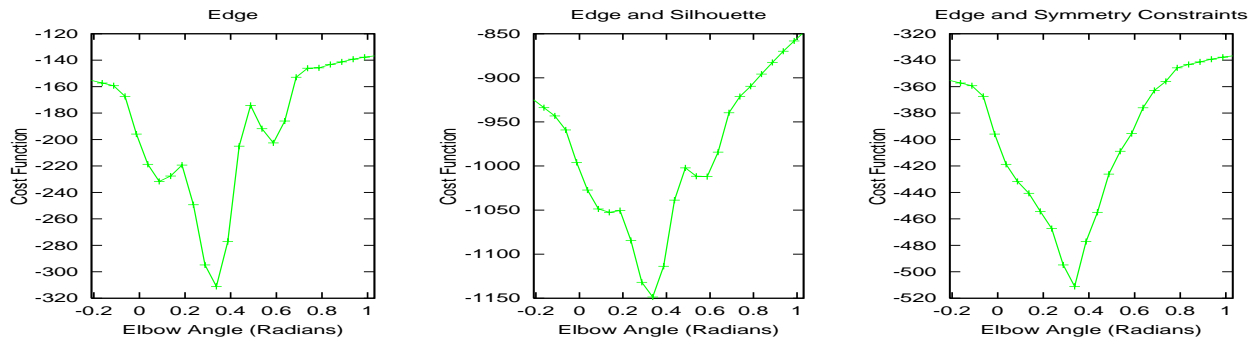


Figure 5: Multi-modality in the cost surface is removed by using edge likelihood with coupled/symmetry constraints

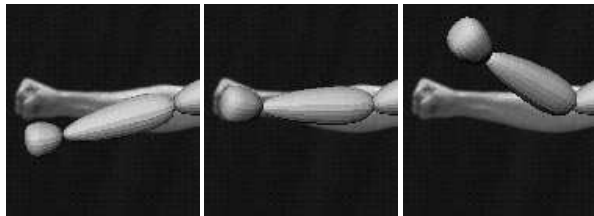


Figure 4: Cost Function Experiment. The elbow joint is varied and the corresponding edge cost is monitored. Multi-modal behavior arises due to incorrect matches during independent edgels search. Matching under model consistency constraints alleviates the problem see Fig. 5

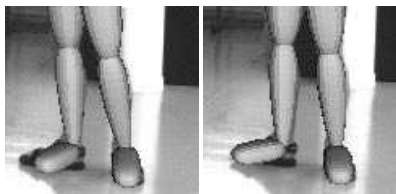


Figure 6: Contour symmetry alleviates data association

modeling and estimation. Motivated by the desire to build cost surfaces whose minima more truly reflect the good configurations in the problem, we proposed two novel likelihood models: (I) a silhouette term based on an attraction/explanation pair and (II) a contour likelihood term that uses robustness and higher-order feature coupling to consistently drive the model-based search process. This allows for significantly better model fits using continuous estimation, eliminates spurious minima and consolidates the true ones, and provides a more comfortable basin of attraction zones and thus pose recovery.

Our future work is aimed towards a more careful analysis of the cost terms' behavior and their singularities and limitations

as well as on the connections between the amount of multi-modality in the cost surface and the amount of search required for a certain model parameter estimation accuracy.

**Acknowledgments** The author thanks Bill Triggs for many helpful discussions on the topics presented in this paper.

## References

- [1] M. Brand. Shadow Puppetry. In *ICCV*, pages 1237–1244, 1999.
- [2] C. Bregler and J. Malik. Tracking People with Twists and Exponential Maps. In *CVPR*, 1998.
- [3] T. Cham and J. Rehg. A Multiple Hypothesis Approach to Figure Tracking. In *CVPR*, volume 2, pages 239–245, 1999.
- [4] Q. Delamarre and O. Faugeras. 3D Articulated Models and Multi-View Tracking with Silhouettes. In *ICCV*, 1999.
- [5] J. Deutscher, A. Blake, and I. Reid. Articulated Body Motion Capture by Annealed Particle Filtering. In *CVPR*, 2000.
- [6] R. Fletcher. Practical Methods of Optimization. In *John Wiley*, 1987.
- [7] D. Gavrilu and L. Davis. 3-D Model Based Tracking of Humans in Action: A Multiview Approach. In *CVPR*, pages 73–80, 1996.
- [8] D. Geiger, A. Gupta, L. Costa, and J. Vlontzos. Dynamic Programming for Detecting, Tracking and Matching Deformable Contours. *PAMI*, 17(3), 1995.
- [9] T. Heap and D. Hogg. Wormholes in Shape Space: Tracking Through Discontinuities Changes in Shape. In *ICCV*, pages 334–349, 1998.
- [10] I. Kakadiaris and D. Metaxas. Model-Based Estimation of 3D Human Motion with Occlusion Prediction Based on Active Multi-Viewpoint Selection. In *CVPR*, pages 81–87, 1996.
- [11] T. L. Liu and D. Geiger. Visual Deconstruction: Reconstructing Articulated Objects. In *Venice Workshop*, 1998.
- [12] P. Perez, A. Blake, and M. Gangnet. JstStream: Probabilistic Contour Extraction with Particles. In *ICCV*, 2001.
- [13] R. Rosales and S. Sclaroff. Inferring Body Pose without Tracking Body Parts. In *CVPR*, pages 721–727, 2000.
- [14] C. Sminchisescu and B. Triggs. Covariance-Scaled Sampling for Monocular 3D Body Tracking. In *CVPR*, 2001.
- [15] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: Recovering 3-D shape and non-rigid motion. *A.I.*, 36(1), 1988.
- [16] K. Toyama and A. Blake. Probabilistic Tracking in a Metric Space. In *ICCV*, 2001.
- [17] S. C. Zhu. Embedding Gestalt Laws in Markov Random Fields. *PAMI*, 21(11), 1999.