

# Human Motion Tracking by Registering an Articulated Surface to 3-D Points and Normals<sup>1</sup>

Radu Horaud (Radu.Horaud@inrialpes.fr)  
Matti Niskanen, Guillaume Dewaele, Edmond Boyer.  
INRIA Grenoble Rhone-Alpes, France

<http://perception.inrialpes.fr/Publications/2009/HNDB09/>

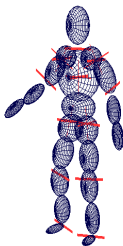
---

<sup>1</sup>IEEE Trans. on PAMI, vol. 31, no. 1, pp 158–164, Jan. 2009

# Introduction

- We address human-motion capture using 3D data gathered with several videos;
- We fit an articulated implicit surface to 3D data, namely points and normals;
- We introduce a new data-to-ellipsoid distance, and
- We derive an expectation-maximization (EM) algorithm that performs the fitting in a reliable and robust maner.

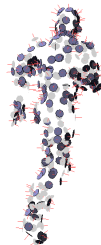
# Outline



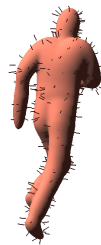
Kinematic model



Implicit surface



3D data



Fitted model

# The articulated model representation

- Articulated implicit surface:

$$f(\mathbf{y}, \mathbf{\Lambda}) = \sum_{p=1}^P \exp\left(-\frac{d_{\mathcal{M}}^2(\mathbf{y}, \mathbf{x}_p(\mathbf{\Lambda}))}{\nu_p^2}\right)$$

- $\mathbf{y} \in \mathcal{S} \Leftrightarrow f(\mathbf{y}) = C$ ,
- $\mathbf{x}_p \in \mathbf{Q}_p$  (an ellipsoid),
- $d_{\mathcal{M}}$  is the data-to-ellipsoid Mahalanobis distance,
- $\mathbf{\Lambda}$  are the kinematic parameters.

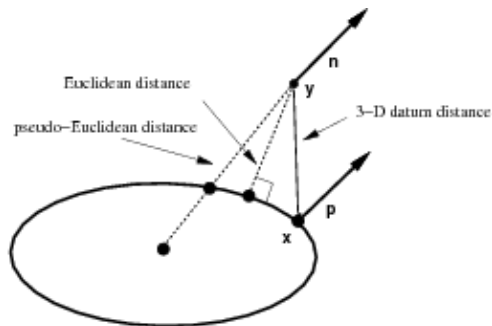


## Problem formulation

The kinematic parameters may be found by minimizing the following non-linear criterion:

$$\mathbf{\Lambda}^* = \arg \min_{\mathbf{\Lambda}} \left( -\nu^2 \sum_{i=1}^I \log \sum_{p=1}^P \exp \left( -\frac{d_{\mathcal{M}}^2(\mathbf{y}_i, \mathbf{x}_{ip}(\mathbf{\Lambda}))}{\nu^2} \right) \right)$$

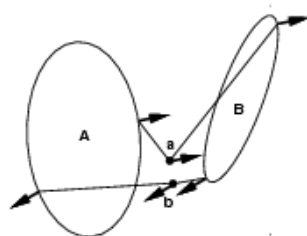
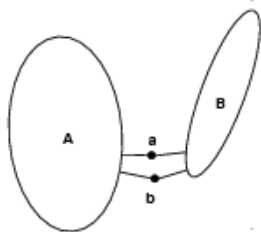
# The data-to-ellipsoid distance



$$d_{\mathcal{M}}^2 = (\mathbf{y}_i - \mathbf{x}_{ip}(\boldsymbol{\Lambda}, \mathbf{n}_i))^{\top} \boldsymbol{\Sigma}_{\mathbf{p}}^{-1} (\mathbf{y}_i - \mathbf{x}_{ip}(\boldsymbol{\Lambda}, \mathbf{n}_i))$$

(See Horaud et al. PAMI'09 for details)

## The advantage of the proposed distance



## Probabilistic *implicit surface* model (1)

- The likelihood of a data point (conditioned by its assignment to an ellipsoid) is drawn from a Gaussian (normal) distribution:

$$P(\mathbf{y}_i | z_i = p) = \mathcal{N}(\mathbf{y}_i | \mathbf{x}_{ip}(\boldsymbol{\Lambda}), \boldsymbol{\Sigma}_p)$$

$$\mathcal{N}(\mathbf{y}_i | \mathbf{x}_{ip}(\boldsymbol{\Lambda}), \boldsymbol{\Sigma}_p) = \frac{1}{(2\pi)^{3/2} |\boldsymbol{\Sigma}_p|^{1/2}} \exp\left(-\frac{1}{2} d_{\mathcal{M}}^2(\mathbf{y}_i, \mathbf{x}_p(\boldsymbol{\Lambda}))\right)$$

- the notation  $z_i = p$  means that the  $i^{th}$  data point (observation) is assigned to the  $p^{th}$  ellipsoid.
- $\mathcal{Z} = \{z_1, \dots, z_I\}$  is the set of *hidden variables* also referred to as *missing data* as opposed to the *observed data*.

## Probabilistic *implicit surface* model (2)

- Apply the sum and product rules of probabilities:

$$P(\mathbf{y}_i) = \sum_{p=1}^P P(\mathbf{y}_i | z_i = p) \underbrace{P(z_i = p)}_{\pi_p}$$

- $\pi_p$  are the **priors** (the **mixing parameters**) with  $\sum_p \pi_p = 1$
- It is assumed that the  $\mathbf{y}_i$  are *independent and identically distributed*, or i.i.d.:

$$P(\mathbf{y}_1, \dots, \mathbf{y}_I) = \prod_{i=1}^I P(\mathbf{y}_i)$$

# The (negative) log-likelihood

- The observed-data log-likelihood:

$$\log P(\mathbf{y}_1, \dots, \mathbf{y}_I) = \sum_{i=1}^I \log P(\mathbf{y}_i)$$

- The negative observed-data log-likelihood becomes:

$$-\sum_{i=1}^I \log \sum_{p=1}^P \frac{\pi_p}{(2\pi)^{3/2} |\boldsymbol{\Sigma}_p|^{1/2}} \exp\left(-\frac{1}{2} d_{\mathcal{M}}^2(\mathbf{y}_i, \mathbf{x}_p(\boldsymbol{\Lambda}))\right)$$

## The two formulas, side by side

- Articulated implicit surface:

$$\arg \min \left( -\nu^2 \sum_{i=1}^I \log \sum_{p=1}^P \exp \left( -\frac{d_{\mathcal{M}}^2(\mathbf{y}_i, \mathbf{x}_{ip}(\boldsymbol{\Lambda}))}{\nu^2} \right) \right)$$

- Observed-data log-likelihood (probabilistic implicit surface):

$$\arg \min \left( -\sum_{i=1}^I \log \sum_{p=1}^P \frac{\pi_p}{(2\pi)^{3/2} |\boldsymbol{\Sigma}_p|^{1/2}} \exp \left( -\frac{1}{2} d_{\mathcal{M}}^2(\mathbf{y}, \mathbf{x}_p(\boldsymbol{\Lambda})) \right) \right)$$

- Both functions are difficult to minimize.

## The expected *complete-data* log-likelihood

- Replace the maximization of the *observed-data log-likelihood* with the maximization of the *expected complete-data log-likelihood* – Expectation-Maximization (EM).
- complete-data means the observed-data  $(\mathbf{y}_1, \dots, \mathbf{y}_I)$  and the missing-data  $(z_1, \dots, z_I)$ :

$$\max E[\log P(\mathbf{y}_1, \dots, \mathbf{y}_I, z_1, \dots, z_I | \mathbf{y}_1, \dots, \mathbf{y}_I)]$$

- EM was introduced by Dempster, Laird, and Rubin in 1977: "Maximum likelihood from incomplete data via the EM algorithm" ([Google scholar](#): **Cited by 18071**)
- We provide an EM algorithm that fits an articulated implicit surface to a set of 3D observed points and normals.



# EM for point-to-surface fitting

- **Initialization.** Provide the kinematic parameters  $\mathbf{\Lambda}^{(q)}$  the covariance matrix  $\mathbf{\Sigma}^{(q)}$  common to all the ellipsoids, and set  $\pi_1^{(q)} = \dots = \pi_P^{(q)} = 1/P$ ;
- **E step.** Evaluate the posterior probabilities  $t_{ip}^{(q)}$  using the current parameter values:

$$t_{ip}^{(q)} = \frac{\pi_p^{(q)} P(\mathbf{y}_i | z_i = p)}{P(\mathbf{y}_i)}$$

- **M step.** Estimate new values for the kinematic parameters  $\mathbf{\Lambda}^{(q+1)}$ :

$$\arg \min_{\mathbf{\Lambda}} \frac{1}{2} \sum_{i=1}^I \sum_{p=1}^P t_{ip}^{(q)} (\mathbf{y}_i - \mathbf{x}_{ip}(\mathbf{\Lambda}))^\top \mathbf{\Sigma}^{(q)-1} (\mathbf{y}_i - \mathbf{x}_{ip}(\mathbf{\Lambda}))$$

## EM (continued)

- **M step (continued).** Update the covariance matrix and the priors:

$$\begin{aligned}\Sigma^{(q+1)} &= \frac{1}{\sum_{i=1}^I \sum_{p=1}^P t_{ip}^{(q)}} \\ &\quad \sum_{i=1}^I \sum_{p=1}^P t_{ip}^{(q)} (\mathbf{y}_i - \mathbf{x}_{ip}(\boldsymbol{\Lambda}^{(q+1)})) (\mathbf{y}_i - \mathbf{x}_{ip}(\boldsymbol{\Lambda}^{(q+1)}))^{\top} \\ \pi_p^{(q+1)} &= \frac{1}{I} \sum_{i=1}^I t_{ip}^{(q)}\end{aligned}$$

- **Maximum likelihood.** Evaluate the observed-data log-likelihood and check for convergence.

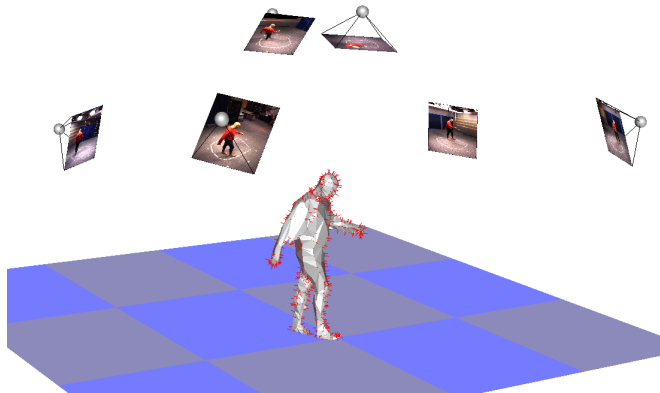
## Practical considerations

- We added a *uniform-noise* component to the Gaussian mixture. This component captures the outliers such that they do not have an impact on the estimated parameters.
- The M-step consists in a non-linear minimization procedure followed by closed-form solutions for the statistical parameters.
- We used a common covariance to all the Gaussian components in the mixture to reduce problems associated with small covariances.
- A good initial solution is necessary.

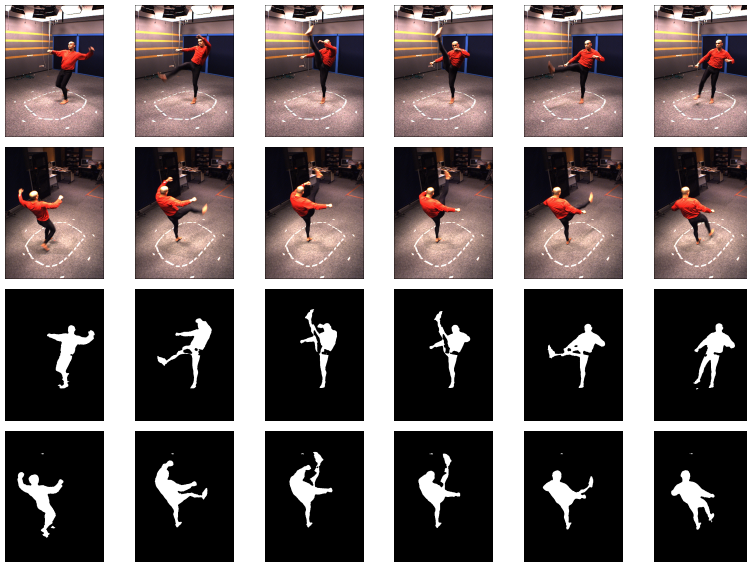
# Results

- We used a multiple-camera setup to gather several videos.
- At each time step we build a visual hull (Franco, Boyer, Lapierre 3DPVT'06, Franco & Boyer PAMI'09).
- We compared the results with the classical distance & model fitting.

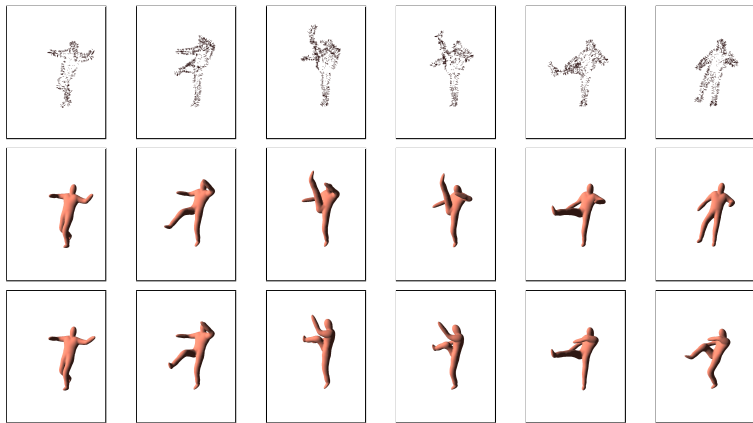
# Camera layout



# Taekwondo (images and silhouettes)



# Taekwondo (3D data, points and normals, points only)



# Conclusions

- A principled probabilistic framework for fitting an articulated implicit surface to 3D points (and normals).
- A modified EM algorithm that estimates simultaneously the kinematic and the statistical parameters of the model.
- Proper initialization is crucial.
- In the future we plan to combine with spectral clustering/matching to solve for many unsolved problems (Diana Mateus, Avinash Sharma, David Knossow, Radu Horaud).