Probabilistic Estimation of Articulated Body Model from Multiview Sequences

Karel Zimmermann* and Tomáš Svoboda**

* Center for Machine Perception, ** Centrum for Applied Cybernetics
Department of Cybernetics, Faculty of Electrical Engineering,
Czech Technical University, Czech Republic

{zimmerk,svoboda}@cmp.felk.cvut.cz

Abstract: An optimization algorithm and statistical description of articulated body model estimation is proposed. The optimization algorithm fits the model into segmented multiview images. The input of our algorithm is a sequence of segmented images captured by several cameras and a structure of the articulated model. The output of the optimization procedure is shape and motion of the articulated model. The optimization runs over all cameras and all images in the sequence. We focus on description and optimization of probability distribution of the model parameters given segmented multiview sequence. We demonstrate the performance of the algorithm on real sequences of walking human.

Introduction

The acquisition of articulated body model, often called motion capture, has numerous applications. Gait analysis, computer animations or ergonomics study are only few examples. The motion capture problem is often solved by using commercially available marker-based systems. A set of markers is attached to important positions on the human body. The 3D coordinates of the markers are computed via magnetic or optical tracking. Such systems are expensive and using markers is uncomfortable for patients and may affect their natural behaviour. Therefore computer vision researches have studied markerless motion capture where no special hardware devices are needed.

A setup for markerless motion capture typically consists of several calibrated [1] cameras encircling a working volume. Static background of the scene is modeled from images of empty scene. Position of human body is found by background subtraction [2] which data allows efficient computation of volumetric model. Motion parameters of the human body are then found by fitting of articulated model to multiview observations or directly to the volumetric model.

Some methods fit model directly to the silhouettes [3, 4]. Then the criterial function describes correspondence of the model projection with silhouettes (e.g. Magnor and Theobalt [5] uses simple xor function). Alternative way is to first reconstruct 3D shape and then fit the model to it. Mikic et al. [6] fit cylindrical model into carved voxel volume. Space carving requires relatively high number of cameras. Therefore Plankers and Fua [7, 8] combine fitting to the silhouettes with stereo reconstruction.

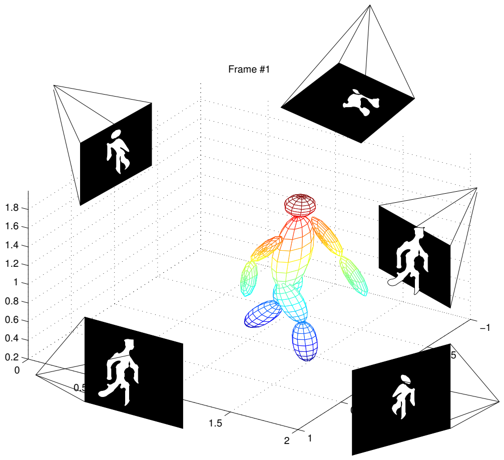

Figure 1: A setup for markerless motion capture consists from several cameras encircling working volume. Articulated ellipsoidal model is used for human body approximation. The parameters are retrieved by optimization where coverage of the silhouettes by model projection is maximized. All parameters are computed fully automatically without any manual intervention. The dimensions are in meters

All mentioned methods have a lot variations in complexity of the model or criterial function, but crucial problem is non-linear high-dimensional minimization. Mikic et al. [6] solve the high-dimensional minimization by hierarchical decomposition (i.e. first fit head then torso and finally limbs), but this, of course, does not assure that global minimum will be found. Urtasun and Fua [9] reduce dimension of search-space by PCA but that limits variety of movement which can be tracked. Deutscher and Reid [10] propose particle filter which is efficient for non-linear searching.

Current state-of-the-art approach achieves good results in human body model tracking but require careful model initialization. The initialization is often done by assigning initial parameters by hand or by requiring a certain pose of the captured human.

Moreover, static parameters evaluation (e.g. length of the arm) requires optimization over the whole sequence at once. In contrast with previous works, we propose an approach with a weak model but we correctly describe

ISSN: 1727-1983 © 2005 IFMBE

decomposition of criterial function over the time to obtain static parameters, close to global minimum over the whole sequence. Hence, our approach needs no manual initialization and exploits all available information since it optimizes over all data.

Human body model and parameters

We decided to use articulated body model created from ellipsoids, see Figure 1. The ellipsoidal model approximates the shape of the real human body and allows very fast projection to the images which is advantageous in the optimization. Each rigid part is represented by one ellipsoid. The structure of the model follows the anatomy of an average human. Movement of a joint causes movement of all succeeding rigid parts. We apply the Hartenberg-Denavit notation for open kinematic chains which is a frequent solution in robotics.

We distinguish two main types of human body model parameters: shape (e.g. length of the arm) and motion (e.g. angle between the upper and the lower arm). The motion parameters are naturally different for each frame of the multiview sequence. However, we assume the shape parameters to be constant. i.e. the person, remain the same throughout the whole sequence. Parameters of the model given multiview sequence $Z = \{Z_1, \ldots, Z_n\}$ are $\theta = \{\mathbf{m}_1, \ldots, \mathbf{m}_n, \mathbf{s}\}$, where n is the number of frames and Z_i is the multiview frame (i.e. set of images from all cameras in time i). In the case of ellipsoidal model, shape parameters are sizes of ellipsoids and motion parameters are mutual positions and orientations.

Let us consider, for the moment, that shape parameters $\mathbf{s}$ are known. Estimation of motion parameters $\mathbf{m}_i$ in frame i is based on posterior probability maximization. The posterior is calculated from projection of the ellipsoidal model to the cameras. The probability of parameters, given images Z_i, is inversely proportional to the sum of distances between border of silhouettes (i.e. segmented images) and projected ellipsoids. The probability is maximized by the standard Gauss-Newton method.

Problem arises when the shape parameters $\mathbf{s}$ are unknown. Clearly, the optimization of all parameters θ over the whole sequence is technically intractable because the number of variables of $p(\theta|Z)$ is proportional to the number of frames in the sequence Z. Therefore we propose an algorithm which finds the maximum of $p(\theta|Z)$ without necessity of optimization over all of the parameters at once. The optimization method is independent on the choice of the model structure (i.e. the derivations are provided without any explicit knowledge of the criterial function). The ellipsoidal model is used only for experimental results and can be simply replaced by more sophisticated model of arbitrary articulated structure.

Statistical framework

In our particular case, we are looking for the most probable shape $\mathbf{s}$ and motion $\mathbf{m}_1, \ldots, \mathbf{m}_n$ parameters, compactly called $\theta = \{\mathbf{m}_1, \ldots, \mathbf{m}_n, \mathbf{s}\}$, given multiview segmented sequence Z of the length n and structure of articulated model. The multiview sequence $Z = \{Z_1, \ldots, Z_n\}$ consists of multiview frames Z_i which are set of images from all of the cameras in time i. The probability of parameters is given by *sequence posterior probability* $p(\theta|Z)$. The sequence posterior optimization over all parameters at once is technically intractable. Therefore we decompose sequence posterior into multiplicative form which is suitable for optimization.

In order to split the optimization task into the particular subtasks we need to accept a few constraints on parameters independence. We expect a human motion to be Markov process. The motion parameters $\mathbf{m}_i$ in frame i are considered to be dependent only on multiview frame Z_i, shape parameters $\mathbf{s}$ and preceding motion parameters $\mathbf{m}_{i-1}$.

Under these constraints, the sequence posterior is decomposed to

$$p(\theta|Z) = p(\mathbf{m}_1|\mathbf{s}, Z_1)p(\mathbf{m}_k, \mathbf{s}|Z_k)\prod_{i \neq k} p(\mathbf{m}_i|\mathbf{m}_{i-1}, \mathbf{s}, Z_i), \tag{1}$$

where k denotes a keyframe (that is arbitrary frame of sequence). The full derivation can be found in Appendix A. The sequence posterior is multiplication of probabilities of parameters in appropriate frames which are called *frame posterior probabilities*. The algorithm can optimize $p(\mathbf{m}_k, \mathbf{s}|Z_k)$ to obtain the optimal value of $\mathbf{s}$ with respect to the observation Z_k. Given the shape parameters we can optimize motion parameters of the whole sequence frame by frame from $p(\mathbf{m}_1|\mathbf{s}, Z_1)$ to $p(\mathbf{m}_n|\mathbf{m}_{n-1}, \mathbf{s}, Z_n)$. These shape parameters are optimal with respect to the frame k, but it is not clear whether these parameters will be optimal for the sequence posterior. Nevertheless, since the frame posterior is the probability of parameters in the given frame, we would expect it to have the maximum near the optimal value of shape parameters with respect to the sequence posterior.

Not all the frames provide the same information about the shape parameters. Two examples of different frames of projection arm-like object to the camera are depicted in Figure 2. The shape parameter $\mathbf{s}$ is in this case the ratio of semi-axes of the two-ellipsoidal model. The first frame does not provide any information about the shape parameter $\mathbf{s}$ because the frame posterior is uniformly distributed. In contrast, different image of the same object allows estimation of the true value of shape parameter $\mathbf{s}$. What is the most informative frame is hard to say in advance. We propose to try the optimization of shape parameters in a several different frames. Given individual hypothesis of shape parameters, we are able to evaluate the sequence posterior and decide which value is optimal. More formal description is provided in the section .

Individual frame posteriors in the (1) are derived in Appendix B. There are two different frame posterior:
with *not given* shape parameters

$$p(\mathbf{m}_i, \mathbf{s}|Z_i) \propto p(Z_i|\mathbf{m}_i, \mathbf{s})p(\mathbf{s}), \tag{2}$$

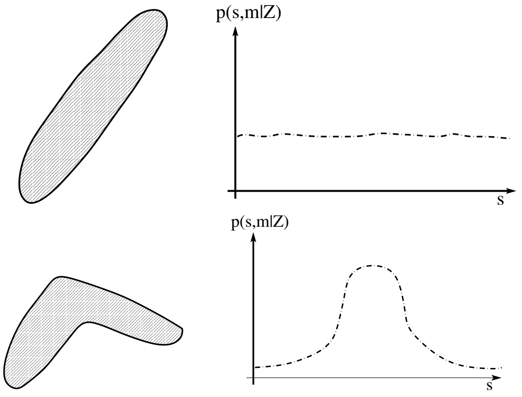

Figure 2: Two views of arm-like object. It consists of two rigid parts. However, the upper view provides no evidence of that fact. This non-usefulness is reflected by the uniform distribution of the shape parameters. The bottom row shows much more useful frame which is reflected by the distribution.

and *given* shape parameters

$$p(\mathbf{m}_i|\mathbf{m}_{i-1},\mathbf{s},Z_i) \propto p(Z_i|\mathbf{m}_i,\mathbf{s})p(\mathbf{m}_i|\mathbf{m}_{i-1}). \qquad (3)$$

If preceding motion parameters are unknown, then the prior $p(\mathbf{m}_i|\mathbf{m}_{i-1})$ has an uniform distribution and equation (3) reduces to the simpler form without this prior. The $p(Z_i|\mathbf{m}_i,\mathbf{s})$, usually called *likelihood*, is proportional to the coverage of segmented images Z_i by model with given parameters $(\mathbf{m}_i,\mathbf{s})$. The likelihood can accommodate arbitrary appearance information like color histogram, edges, etc.

In this work, however, we do not use any appearance information and likelihood is proportional to the distance between border of silhouettes (segmented images) and projected ellipsoidal model [1]. The $p(\mathbf{m}_i|\mathbf{m}_{i-1})$ and $p(\mathbf{s})$ are prior probabilities of motion $\mathbf{m}_i$ and shape $\mathbf{s}$ parameters, respectively. The prior of motion parameters given preceding motion parameters combines temporal coherence constraints with natural motion limitations. The shape prior, which represents the probability of the shape, is considered to have a Gaussian distribution with the mean and covariance matrix proportional to the natural shapes.

Optimization

The sequence posterior is expressed as multiplication of frame posterior probabilities in (1) where only one of them (no matter which) is not given shape parameters in advance. Theoretically, we could obtain shape parameters by optimization $p(\mathbf{m}_k,\mathbf{s}|Z_k)$ in arbitrary frame k. As argued in preceding section however, the approximation of true value is successful only in the most informative frames.

The shape parameter optimization is performed by Algorithm 1. The vector of shape parameters $\mathbf{s}$ is called hypothesis. Given the hypothesis, motion parameters are evaluated frame by frame by maximizing the frame posteriors. Then, probability of the hypothesis is calculated by (1).

First, the algorithm chooses the set of keyframes, frames which expects to be the most informative. The selection follows the expected type of motion and the sampling frequency (frames per second). The keyframes should (sparsely) cover at least one period of the motion. The first shape hypothesis is shape which maximize the prior $p(\mathbf{s})$. Next shape hypothesis are obtained by optimization in the keyframes over the shape and motion parameters.

Second, the motion parameters, given each of hypothesized shape parameters, are evaluated. Finally, sequence posterior of each hypothesis (i.e. shape and appropriate motion parameters) is calculated and the most probable hypothesis is selected.

Algorithm 1 - the sequence posterior maximization.

```
(1) $\mathbf{K}$ is set of keyframe indexes, $\mathbf{H}$ is the set of shape parameters hypothesis.
(2) set the shape parameters to the mean of prior $p(\mathbf{s})$ (i.e the most apriori probable values).
(3) for each keyframe $i \in \mathbf{K}$:
      • optimize shape and motion parameters $(\mathbf{m}^*,\mathbf{s}^*) = \arg\max p(\mathbf{m}_i,\mathbf{s}|Z_i)$ (equation (2))
      • save the $\mathbf{s}^*$ as hypothesis $\mathbf{H} = \mathbf{H} \cup \mathbf{s}^*$.
(4) for each shape hypothesis $\mathbf{s}_j \in \mathbf{H}$ and each frame $i$ do:
      • optimize only motion parameters $\mathbf{m}^* = \arg\max p(\mathbf{m}_i|\mathbf{m}_{i-1},\mathbf{s},Z_i)$ (equation (3)).
(5) evaluate sequence posterior (equation (1)) for each $\mathbf{s}_j \in \mathbf{H}$ and appropriate motion parameters and choose the most probable hypothesis. $k$ is number of appropriate keyframe.
```

To make the proposed algorithm working we must accept some constraints which assure that we, at least asymptotically, reach the true value of shape parameters. We have not yet mentioned any constraints on the character of the frame posterior probability. Let us denote $\bar{\mathbf{s}}$ true (unknown) shape parameters which maximize sequence posterior

$$\bar{\mathbf{s}} = \arg\max p(\theta|Z).$$

Let us denote $\mathbf{s}_k^*$ the shape parameters which maximize the frame posterior in a frame k. Clearly, if the $\mathbf{s}_k^* = \bar{\mathbf{s}}$ then the maximization of the frame posterior in arbitrary frame provides the true value of shape parameters, but this is too hard constraint. Nevertheless, since the frame posterior is the probability of parameters in the given frame, we should expect it to have the maximum near the $\bar{\mathbf{s}}$.

By the maximization of the second term $p(\mathbf{m}_k,\mathbf{s}|Z_k)$ of (1) we can find the maximum of frame posterior probability $\mathbf{s}_k^*$. This value is not equal to the optimal value of $\bar{\mathbf{s}}$ but we expect to be close to it. By the optimization of the

[1] More details about the model can be found in Appendix C

different frames we obtain different values $\mathbf{s}_k^*$. We can evaluate the sequence posteriors given these two different shape parameters by the optimization in the remaining frames. Now we express the probability that after the maximization of shape parameters in K different frames we obtain at least one vector of shape parameters which is closer to the $\bar{\mathbf{s}}$ then ε.

More formally, the maxima of frame posteriors are considered to have a Gaussian distribution with the mean $\bar{\mathbf{s}}$ and a covariance σ. The covariance, which corresponds to the model and likelihood selection, is considered to be as small as possible. Then

$$e = \bar{\mathbf{s}} - \mathbf{s}_k^* = \mathbf{N}(0,\sigma)$$

is of Gaussian distribution too. Note, that we can find ML estimation of σ from different values of $\mathbf{s}_k^*$.

The probability that the maximum $\mathbf{s}_k^*$ of the frame posterior in frame k is closer then ε to the true value $\bar{\mathbf{s}}$ is

$$p_{\varepsilon,\sigma} = \int_{-\varepsilon}^{\varepsilon} \mathbf{N}(e,0,\sigma)de.$$

If we find the $\mathbf{s}_k^*$ in the K frames independently then the probability that the maximum $\mathbf{s}_k^*$ of the frame posterior in at least one of these K frames is closer then ε to the true value $\bar{\mathbf{s}}$ is

$$p_{\varepsilon,\sigma}(K) = 1 - (1 - p_{\varepsilon,\sigma})^K.$$

This function rapidly goes to the 1 in K for reasonable values of ε and σ. In the end of the Algorithm 1 we obtain K different values of s_k^* $k \in \mathbf{K}$ which are used to find a ML estimation of σ. One of this shape parameters provide maximal value of sequence posterior. The $p_{\varepsilon,\sigma}(K)$ is the probability that this shape parameters are closer than ε to the true value of $\bar{\mathbf{s}}$.

Experiments

Simulated data

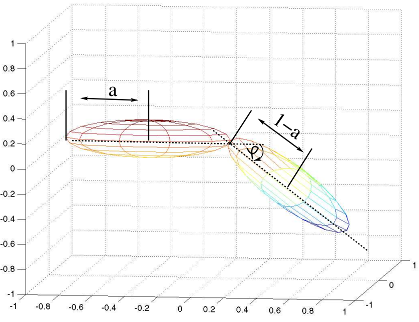

Figure 3: Human arm model. $\theta_i = (\varphi_i, a)$

In the first experiment we clarify basic principle in simple example shown in Figure 3. We fit a human arm model (i.e. open kinematic chain of two ellipsoids) into an observation sequence (6 frames). We decided for parameters $\theta_i = (\varphi_i, a)$, where φ_i (motion parameter) is an angle between two main axes and a (shape parameter) is the length of the main semi-axis of the first ellipsoid. Length of the main semi-axis of second ellipsoid is $1-a$ and all of the other parameters are fixed. The unknown parameters of the model are $a = 1-a = 0.5$, and motion parameters changes frame by frame $\varphi_1 = 0\,\mathrm{rad}$, $\varphi_2 = 0.1\,\mathrm{rad}$, $\varphi_2 = 0.2\,\mathrm{rad}, \ldots$

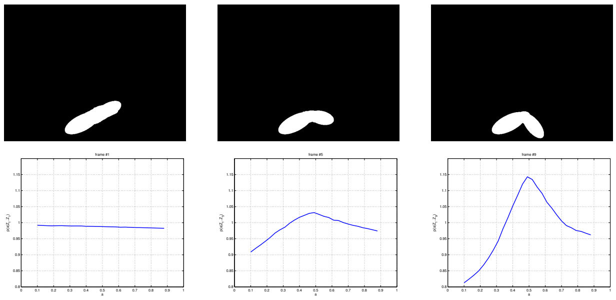

Figure 4: First row shows observation sequence (silhouettes) in the frames 1,4,6. Second row shows appropriate sequence probabilities $p(a,\varphi_1^*|Z_1), p(a,\varphi_1^* \ldots \varphi_4^*|Z_1 \ldots Z_4), p(a,\varphi_1^* \ldots \varphi_6^*|Z_1 \ldots Z_6)$.

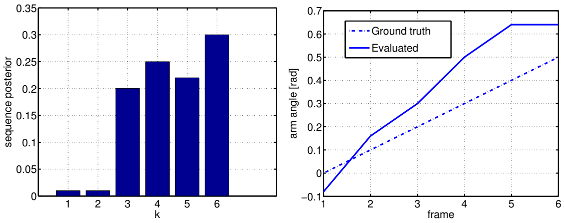

Figure 5: Left: Visualization of probability of different hypothesis. Right: φ values associated with most probable explanation and ground truth.

According to **Algorithm 1**, we obtain shape parameter independently in each keyframe (note, that all frames are keyframes in this simple example). Distribution of shape parameter a and input sequence are depicted in Figure 4. Shape parameters computed for different frames are, of course, different. Each instance of shape parameters is called hypothesis k. Given the hypothesis, we computes motion parameters of the model for whole sequence and probability of this hypothesis. Values of normalizated sequence probabilities for each hypothesis are depicted in Figure 5a.

We can see, that first frames do not provide any information about the shape parameters because posterior is equally distributed (see Figure 4). Thus, we obtain $a = 0.18$ from the first frame and corresponding incorrect vector of motion parameters . Since the third frame shape starts appear and correct shape and motion parameters are calculated. True and the most probable values of the angle φ^* are in Figure 5b. The most probable shape parameter $a = 0.53$. The approximation is inaccurate due to the simulated noise.

Real sequence

We used multiview segmented sequence [2] of human gait captured by seven cameras. The length of sequence is about 200 frames. We select each twentieth frame as keyframe (i.e. 10 keyframes were used). Parameters estimated by our method matched well with the ground truth which was acquired by a complicated, manually intervened, process. All of the shape parameters (length of limbs and sizes of body) were calculated correctly. Motion parameters were computed correctly up to frames about number 120 (see Figure 6). The ellipsoidal model is too weak to correctly distinguish arm angles when arms are close to body and therefore almost invisible in segmented views. It should be noted however, that this is problem of the weak model not the method itself. Correct motion parameters produce higher (worse) value of the criterial function. After the problematic frame, arms are distinguishable from the body the motion parameters returns to the correct values.

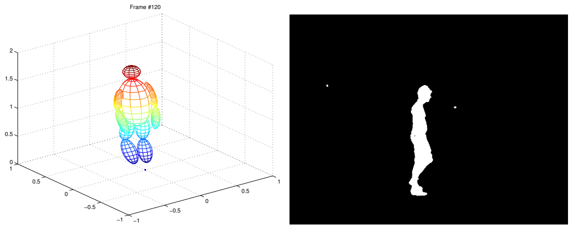

Figure 6: Frame 120 - incorrect motion parameters of arm due to coverage of arms by body. Human arm bends to the opposite side to cover inaccuracy of the model.

Conclusions

We proposed a method of human body model fitting into segmented multiview data, which optimize both shape and motion parameters over the whole sequence. Main contribution is the probabilistic description which, under reasonable constraint, provides solution near to global optimum. Moreover, we are able to quantify probability that we have actually reach the global minimum.

Acknowledgment

The authors thank Lars Muendermann from Stanford University for providing the data. Karel Zimmermann was supported by The Czech Academy of Sciences under project 1ET101210407. Tomáš Svoboda was supported by The Czech Ministry of Education under project 1M6840770004 and by The STINT under project Dur IG2003-2 062 and by The Austrian Ministry of Education under project CONEX GZ 45.535.

Appendix A

We will show how to decompose sequence probability into multiplicative form which is usable for optimization. The Bayes rule application on sequence posterior probability $p(\theta|Z)$ provides

$$p(\mathbf{m}_1,\ldots,\mathbf{m}_n,\mathbf{s}|Z_1,\ldots,Z_n) =$$

$$p(\mathbf{m}_k,\mathbf{s}|Z_1,\ldots,Z_n)p(\mathbf{m}_1,\ldots,\mathbf{m}_{k-1},\mathbf{m}_{k+1},\ldots,\mathbf{m}_n|\mathbf{m}_k,\mathbf{s},Z_1,\ldots,Z_n).$$

The motion and static parameters $(\mathbf{m}_k,\mathbf{s})$ in the frame k are dependent only on the multiview image Z_k. Therefore the first term is reduced to

$$p(\mathbf{m}_k,\mathbf{s}|Z_1,\ldots,Z_n) = p(\mathbf{m}_k,\mathbf{s}|Z_k).$$

The Bayes rule is similarly applied for the second term to obtain similar decomposition

$$p(\mathbf{m}_1,\ldots,\mathbf{m}_{k-1},\mathbf{m}_{k+1},\ldots,\mathbf{m}_n|\mathbf{m}_k,\mathbf{s},Z_1,\ldots,Z_n) =$$

$$p(\mathbf{m}_1|\mathbf{s},Z_1)p(\mathbf{m}_2,\ldots,\mathbf{m}_{k-1},\mathbf{m}_{k+1},\ldots,\mathbf{m}_n|\mathbf{m}_k,\mathbf{s},Z_1,\ldots,Z_n)$$

This expression again consists of two terms where second is decomposed by similar chain application of Bayes rule to the form

$$p(\mathbf{m}_2,\ldots,\mathbf{m}_{k-1},\mathbf{m}_{k+1},\ldots,\mathbf{m}_n|\mathbf{m}_k,\mathbf{s},Z_1,\ldots,Z_n) =$$

$$\prod_{i\neq k} p(\mathbf{m}_i|\mathbf{m}_{i-1},\mathbf{s},Z_i).$$

Substituting these results to the first equation we obtained wanted decomposition of the sequence posterior

$$p(\theta|Z) = p(\mathbf{m}_1|\mathbf{s},Z_1)p(\mathbf{m}_k,\mathbf{s}|Z_k)\prod_{i\neq k} p(\mathbf{m}_i|\mathbf{m}_{i-1},\mathbf{s},Z_i).$$

Appendix B

In preceding Appendix we express sequence probability as multiplication of frame posterior probability. Here, we derive how to compute different posterior probabilities.

The frame posterior probability is derived from generalized Bayesian rule. The joint probability $p(\mathbf{m}_i,\mathbf{m}_{i-1},Z_i,\mathbf{s})$ can be expressed in different forms based on different order of decompositions

$$p(\mathbf{m}_i,\mathbf{m}_{i-1},Z_i,\mathbf{s}) = p(\mathbf{m}_i,\mathbf{s}|\mathbf{m}_{i-1},Z_i)p(\mathbf{m}_{i-1}|Z_i)p(Z_i)$$

$$p(\mathbf{m}_i,\mathbf{m}_{i-1},Z_i,\mathbf{s}) = p(Z_i|\mathbf{m}_i,\mathbf{m}_{i-1},\mathbf{s})p(\mathbf{m}_i,\mathbf{s}|\mathbf{m}_{i-1})p(\mathbf{m}_{i-1}).$$

Therefore

$$p(\mathbf{m}_i,\mathbf{s}|\mathbf{m}_{i-1},Z_i) = \frac{p(Z_i,\mathbf{s}|\mathbf{m}_i,\mathbf{m}_{i-1},\mathbf{s})p(\mathbf{m}_i,\mathbf{s}|\mathbf{m}_{i-1})p(\mathbf{m}_{i-1})}{p(\mathbf{m}_{i-1}|Z_i)p(Z_i)}.$$

We do not model any prior knowledge about the probability of multiview image Z_i and expect it uniformly distributed, i.e. $p(Z_i) = const$. The motion parameters $\mathbf{m}_{i-1}$ in the frame $(i-1)$ are independent on the image Z_i in the frame i, therefore $p(\mathbf{m}_{i-1}|Z_i) = p(\mathbf{m}_{i-1})$. There are

[2] Data have been provided by Lars Muendermann from Stanford University

no explicit knowledge about the shape and motion parameters and therefore are considered to be independent too $p(\mathbf{m}_i,\mathbf{s}|\mathbf{m}_{i-1}) = p(\mathbf{m}_i|\mathbf{m}_{i-1})p(\mathbf{s})$. By substituting these expressions and ignoring constant we obtain exactly

$$p(\mathbf{m}_i,\mathbf{s}|\mathbf{m}_{i-1},Z_i) \propto p(Z_i|\mathbf{m}_i,\mathbf{s})p(\mathbf{m}_i|\mathbf{m}_{i-1})p(\mathbf{s}).$$

Similarly, from different forms of joint probability $p(\mathbf{m}_i,\mathbf{m}_{i-1},Z_i,\mathbf{s})$, we derive motion $\mathbf{m}_i$ and shape $\mathbf{s}$ joint probability given preceding motion $\mathbf{m}_{i-1}$ and images Z_i. Comparison of equations

$$p(\mathbf{m}_i,\mathbf{m}_{i-1},Z_i,\mathbf{s}) = p(\mathbf{m}_i|\mathbf{m}_{i-1},Z_i,\mathbf{s})p(\mathbf{s}|\mathbf{m}_{i-1},Z_i)p(\mathbf{m}_{i-1},Z_i),$$

$$p(\mathbf{m}_i,\mathbf{m}_{i-1},Z_i,\mathbf{s}) = p(Z_i|\mathbf{s},\mathbf{m}_i,\mathbf{m}_{i-1})p(\mathbf{m}_i|\mathbf{s},\mathbf{m}_{i-1})p(\mathbf{s},\mathbf{m}_{i-1})$$

provides

$$p(\mathbf{m}_i|\mathbf{m}_{i-1},Z_i,\mathbf{s}) = \frac{p(Z_i|\mathbf{s},\mathbf{m}_i)p(\mathbf{m}_i|\mathbf{m}_{i-1})p(\mathbf{s},\mathbf{m}_{i-1})}{p(\mathbf{s}|\mathbf{m}_{i-1},Z_i)p(\mathbf{m}_{i-1},Z_i)}.$$

Considering independences mentioned above,

$$p(\mathbf{m}_i|\mathbf{m}_{i-1},Z_i,\mathbf{s}) = \frac{p(Z_i|\mathbf{s},\mathbf{m}_i)p(\mathbf{m}_i|\mathbf{m}_{i-1})p(\mathbf{s})p(\mathbf{m}_{i-1})}{p(\mathbf{s})p(\mathbf{m}_{i-1})p(Z_i)}.$$

Simplification of this expression provides

$$p(\mathbf{m}_i|\mathbf{m}_{i-1},\mathbf{s},Z_i) \propto p(Z_i|\mathbf{m}_i,\mathbf{s})p(\mathbf{m}_i|\mathbf{m}_{i-1}).$$

Appendix C

The frame posterior probability (2), (3) consists of likelihood and priors. The priors are probabilities expressing knowledge given in advance and are considered to have a simple distribution (i.e. uniform or Gaussian). The likelihood express the probability that the image was induced by parameters $(\mathbf{m}_i,\mathbf{s})$. Such probability can be expressed in many ways. We chose to make it proportional to the coverage of borders of silhouettes by model projection in the individual cameras. The model projection and borders of silhouettes are depicted in the Figure 7. The sum of distances of pixels from the model projection over all cameras is inversely proportional to the frame posterior probability. The distance of the pixel from the model projection is considered to be the distance of the pixel from the nearest ellipse.

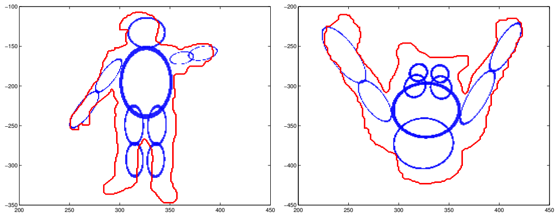

Figure 7: The projection of ellipsoidal model to the individual cameras (blue color) and the border of segmented image (red color)

The Euclidean distance between point $\mathbf{x}$ and ellipse requires solving of quartic equation which may have up four solutions, requiring the one with the minimum distance to be determined [11]. To avoid the complexity of evaluating the true Euclidean distance, we proposed an approximation measure [12].

References

[1] T. Svoboda, D. Martinec, and T. Pajdla. A convenient multi-camera self-calibration for virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 14(4), August 2005.

[2] C. Stauffer and W.E.L Grimson. Adaptive background mixture models for real-time tracking. In *IEEE Computer Vision and Pattern Recognition*, pages 2246–2252, June 1999.

[3] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel. Free-viewpoint video of human actors. *ACM Transaction on Computer Graphics*, 22(3):569–577, July 2003.

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

[5] M. Magnor and C. Theobalt. Model-based analysis of multi-video data. *IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 41–45, 2004.

[6] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman. Human body model aquisition and tracking using voxel data. *International Journal of Computer Vision*, 3(53):199–223, 2003.

[7] R. Plankers and P. Fua. Articulated soft objects for multi-view shape and motion capture. *Pattern Analysis and Machine Inteligence*, 25(9):1182–1187, 2003.

[8] R. Plankers and P. Fua. Articulated soft objects for video-based body modeling. In Springer, editor, *Proc. International Conference on Computer Vision*, pages 394–401, Springer-Verlag Berlin Heidelberg, 2001.

[9] R. Urtasun and P. Fua. 3d tracking for gait characterization and recognition. In Springer, editor, *Proc. European Conference on Computer Vision*, pages 17–22, Springer-Verlag Berlin Heidelberg, may 2004. Springer.

[10] J. Deutscher and I. Reid. Articulated body motion capture by stochastic search. *International Journal of Computer Vision*, 61(2):185–205, 2005.

[11] P. L. Rosin. Analysing error of fit functions for ellipses. *Pattern Recognition Letters*, 17(14):1461–1470, 1996.

[12] Karel Zimmermann and Tomáš Svoboda. Approximation of Euclidean distance between point from ellipse. Research Report CTU–CMP–2005–23, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, August 2005.