Pose Estimating the Human Arm Using Kinematics and the Sequential Monte Carlo Framework

Thomas Moeslund

1. Introduction

Human Motion Capture (MoCap) has for several years been an increasing activity around the world. The primary reason being the number of applications where MoCap is a necessity. Most well know is perhaps the application areas of Human Computer Interaction (HCI), surveillance, diagnostics of orthopaedic patients, analysis of athletes, and computer graphics animations such as the motion of the dinosaurs in ”Jurassic Park” and Gollum in ”Lord of the Rings”. Different MoCap technologies exist but marker-free computer vision-based MoCap is of special interest as it can provide a ”touch-free” and one-sensor (monocular) technology, see (Moeslund & Granum, 2001) for a survey.

Computer visions-based MoCap systems use different number of cameras, where, generally speaking, the fewer cameras the more complex the task. When more cameras are applied one of two overall techniques are applied. Either the position of the individual body parts are found in each image and combined into 3D-positions using triangulation (Azoz *et al.*, 1998); (Wren *et al.*, 2000), or a disparity map of the image is produced and compared with a 3D model of the human (Fua *et al.*, 1998);(Plankers *et al.*, 1999). A hybrid approach is to have multiple cameras but only use the one with the 'best view' to process the current image (Ong & Gong, 1999).

When just one camera is present a geometric model of the human is used to solve the ill-posed problem of estimating a 3D pose based on a sequence of 2D images. The model is either used directly in the estimation or it is used indirectly. By the latter is meant that the model guides the estimation process, as opposed to the former where the model is an integrated part of the estimation process. For example, in the work by (Segawa *et al.*, 2000) a particular pose of an arm is estimated by filtering past and future image measurements of the arm. The filtering is done using an Extended Kalman Filter wherein the joint angle limits of the model are incorporated. The model is therefore not used directly but rather indirectly to constrain the possible solutions.

In the case of direct model use, the model is synthesised into the image where it is compared to the image data. This process is known as the analysis-by-synthesis (AbS) approach. The type of data used in the matching differ between systems, but usually it is: edges (Sminchisescu, 2002);(Wu *et al.*, 2003), texture (Sidenbladh *et al.*, 2000);(Ben-Arie *et al.*, 2002), contours (Ong & Gong, 1999);(Delamarre & Faugeras, 2001), silhouettes (Moeslund & Granum, 2000);(Ogaki *et al.*, 2001), or motion (Bregler & Malik, 1998);(Howe *et al.*, 2000).

Source: Cutting Edge Robotics, ISBN 3-86611-038-3, pp. 784, ARS/plV, Germany, July 2005 Edited by: Kordic, V.; Lazinica, A. & Merdan, M.

The framework used to decide which synthesised data to match with the current image data also differ. Usually a very high number of different model poses exist and an exhaustive matching is seldom realistic. Different approaches are therefore applied. One approach is to synthesise just one model pose and let the difference between the synthesised data and the image measurements be used as an error signal to correct the state of the model (Goncalves *et al.*, 1995); (Wren *et al.*, 2000). An excellent framework for this type of approach is the Kalman Filter.
Another approach is to formulate the matching as a function of the different parameters in the model. This function is then optimised, i.e., different model poses are synthesised until the synthesised data and image data is close with respect to some metric. Due to the high dimensionality of the problem an analytic solution is not possible and a numeric iterative approach is generally used (Gavrila & Davis, 1996);(Wachter & Nagel 1999). Within the last five years or so stochastic approaches have been the preferred ways of handling the dimensionality problem. Especially, the Sequential Monte Carlo (SMC) framework has been used successfully (Isard & Blake, 1998);(Deutscher *et al.*, 2000);(Mitchelson & Hilton, 2003);(Lee & Cohen, 2004). The SMC framework is basically a Bayesian approach where the current state of the model is represented by the posterior information inferred from predicted the most likely states from the previous frame and validating them using the current image measurements.
Independent of how the pose estimation problem is formulated the AbS-approach results in a huge solution-space. Therefore kinematic constraints are often applied in order to prune the solution-space, e.g., the bending of the elbow is between 0 and 145^{o}. This may be used directly to partition the solution-space into legal and illegal regions, as in (Segawa *et al.*, 2000), or the constraints may be defined as forces acting on an unconstrained state phase (Wren *et al.*, 2000). The fact that two human body parts cannot pass through each other also introduces constraints. Another approach to reduce the number of possible model poses is to assume a known motion pattern - especially cyclic motion such as walking and running. In the work by (Rohr, 1997) gait parallel to the image plane is considered. Using a cyclic motion model of gait all pose parameters are estimated by just one parameter, which specifies the current phase of the cycle. This is perhaps the most efficient pruning ever applied! (Ong & Gong, 1999) map training data into the solution-space and use PCA to find a linear subspace where the training data can be compactly represented without loosing too much information. (Pavlovic *et al.*, 1999) take this idea a step farther by learning the possible or rather likely trajectories in the state space from training data, i.e., dynamic models are learned.

1.1 The Content of this Paper

From a HCI point of view the primary interest regarding MoCap is a reliable way of pose estimating the arms over time. The focus of this work is therefore on pose estimating a human arm using monocular computer vision. The approach we take is twofold. First of all we want to apply the image measurements from the current frame in order to 1) derive a more compact representation of the solution-space and 2) improve the SMC framework. Secondly, we want to exploit the kinematic constraints more thoroughly than what is usually the case. The hypothesis behind the approach is that the fewer possible solutions (the result of a compact solution-space and thorough constraints) and the better the search (SMC) through these, the higher the likelihood of actually finding the correct pose in a particular frame.

The structure of the paper is as follows. In section 2 a new way of representing the human arm is presented. In section 3 this model is pruned by exploiting the kinematic constraints. In section 4 the model of the arm is applied in a SMC framework in order to pose estimate an arm. In section 5 results are presented and in section 6 a conclusion is given.

2. Modelling the Arm

When modelling the pose of the arm it is necessary to understand the anatomic features, which controls the movement of the arm, hence the bones and the joints connecting them. A complete description of the interacting between the joints and bones and the DoF this results in, leads to a comprehensive model that again results in a huge solution-space. This is not desirable and we therefore focus on the large-scale motion of the arm meaning that only the most significant DoF in the arm are modelled. Furthermore, we assume that a geometric ideal joint can model each anatomic joint. Another consequence of the focus on large-scale motion is that we assume the hand to be a part of the lower arm.

Overall the arm consists of the upper- and lower arm; those lengths are assumed known as A_u and A_l, respectively. The length of the lower arm is defined as the distance from the elbow to the centre of the hand. As for the rest of the movable rigid entities in the human body, the arm consists of bones that are moved by muscles following the design of the joints connecting the bones. The ligaments, muscles, joints, and other bones impose the limits of the movements, but in this work we simply refer to the limits of a joint independent of the origin.

The lower arm consists of two bones; the ulna and the radius (Morrey, 1985), see figure 2. They are connected to each other with a pivot joint at the elbow and with a pivot joint at the wrist. These two joints allow the hand to rotate around the long axis of the lower arm by rotating the radius around the ulna. The pivot joint in the elbow is part of the elbow joint that also connects the two bones with the upper arm bone, the humerus. The ulna is connected in a hinge joint and the radius in a ball-and-socket joint. Overall the primary DoF at the elbow is modelled very well using one hinge joint and one pivot joint even though the elbow motion is more complex (An & Morrey, 1985). Since we ignore the motion of the hand we can ignore the pivot rotations of the radius around the ulna, hence the DoF in the elbow is modelled by one hinge joint.

The upper arm consists of one bone, the humerus, and is connected to the shoulder in the gleno-humeral joint (GH-joint), see figure 2. Even though the GH-joint contains two sliding DoF (Dvir & Berme, 1978) it is well modelled as a ball-and-socket joint since the surfaces of the joint is more than 99% spherical (Soslowsky *et al.*, 1999).

The "socket" part of the joint is a part of the shoulder and called the glenoid. Its motion with respect to the torso, or more precisely the thorax (Zatsiorsky, 1998), originates from the complex structure of the shoulder, known as the shoulder complex. The shoulder complex provides mobility beyond any other joint in the entire human body (Zatsiorsky, 1998). The shoulder complex contains up to 11 DoF (Maurel, 1998). However, since the mobility of the shoulder complex can be described as a closed kinematic chain these 11 DoF are not independent and in fact the relative pose of the glenoid is well described by only four independent parameters (Zatsiorsky, 1998).

To model the shoulder complex requires a comprehensive biomechanical model based on knowledge of the bones, joints, ligament, muscles, and their interactions. Such models have been developed, see e.g., (Engin & Tumer, 1989);(Hogfors, 1991); (Maurel, 1998).

We can however not use such models since they contain too many details, hence too many parameters. However, by analysing the outcome of advanced biomechanical models we

observe that the primary motion of the glenoid with respect to the torso, hence the four parameters, is: rotations around the z- and y-axes, and vertical and horizontal displacements along the y- and x-axes, see figure 1 for a definition of the axes. The rotations can be governed by the GH-joint by increasing its ranges accordingly whereas two prismatic joints can model the two displacements, each having one DoF.
Altogether six DoF are needed to model the primary DoF of the arm and the shoulder complex. As the prismatic joints have significantly less effect on the pose of the arm compared to the elbow and GH-joints we (for now) ignore the prismatic joints and focus on the remaining four primary DoF.

2.1 Modelling the four DOFs of the Arm

A number of different ways of modelling the four DOF in the arm exist (Moeslund, 2003). The most common model is the one shown in figure 1 where four Euler angles are applied. Since we aim at a compact state-space we derive a new model inspired by the screw axis model.
The parameters in the screw axis model do not directly relate to the anatomic joints. Nevertheless the model has the same ability to represent the different arm configurations as for example the Euler angle model has. The representation is based on Chasles' theorem (Zatsiorsky, 1998) that loosely states that a transformation between two coordinate systems can be expressed as a rotation around an axis, called the screw axis (or helical axis), and a translation parallel to the screw axis. In the context of modelling the human arm the screw axis is defined as the vector spanned by the shoulder and the hand. The position of the elbow is defined as a rotation of an initial elbow position around the screw axis. Since the length of the upper and lower arm are fixed no translation is required parallel to the screw axis and the perpendicular distance from the elbow to the screw axis is independent on the rotation and can be calculated without adding additional parameters. Altogether the representation requires four parameters; three for the position of the hand and one for the rotation around the screw axis, α. In figure 1 the parameters are illustrated.

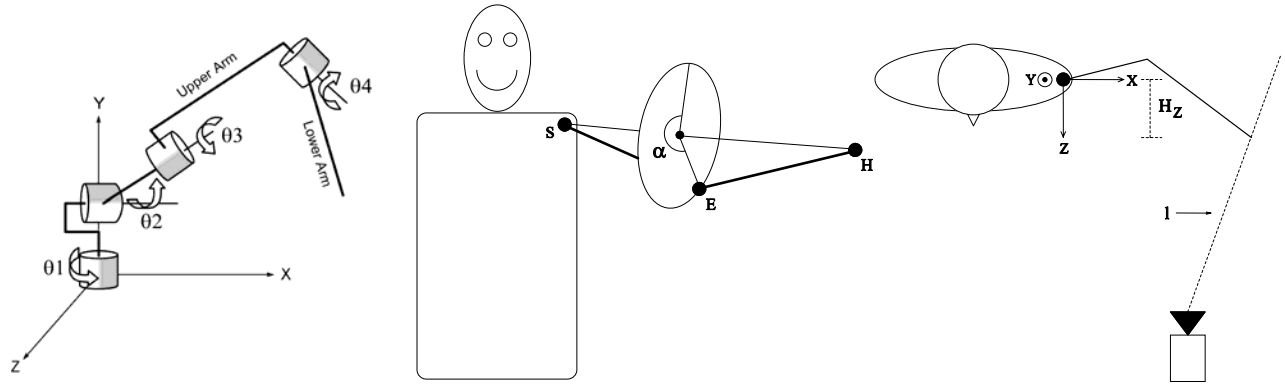


Figure 1. Different arm representations. **Left**: Four Euler angles. **Middle**: Screw axis model. **Right**: Local screw axis model

We now combine the screw axis model with image data in order to obtain a more compact representation of the arm. Colour information is used to segment the hand in an image. Combining this with the camera parameters obtained during calibration the position of the hand in an image can be mapped to a line in space:

$$\vec{H}(t) = \vec{P} + t \cdot \vec{D} \tag{1}$$

where $\vec{P}$ is a point on the line, e.g., the focal point of the camera, and $\vec{D}$ is the unit direction vector of the line. This means that for each value of H_Z the two other components H_X and H_Y are uniquely determined. Applying this to the screw axis representation we obtain a "local screw axis model" utilising only two parameters, α and H_Z, to model the pose of the arm. For each new image a unique instance of the solution-space exist, hence the name "local". In this compact model α is bounded by one circle-sweep [0^o , 360^o], while H_Z is bounded by $\pm$ the total length of the arm, A_u+A_l. Using only two parameters to model the arm is indeed a compact representation. Furthermore, having only two parameters also allows for a visualisation of the result when comparing the synthesised data with the image data, hence the solution-space can be directly visualised - a very powerful characteristic!

2.2 Eliminating the Effect of the Prismatic Joints

Previously we suggested using six parameters to model the arm: two for the shoulder complex, three for the GH-joint, and one for the elbow joint. Next we suggested ignoring the prismatic joints by arguing that they have significantly less effect on the pose of the arm compared to the elbow and GH-joints. Even though this is true the prismatic joints should not be ignored altogether. In this section we therefore revise the prismatic joints and include them in our local screw axis model.

Concretely we seek to eliminate the effect of the prismatic joints altogether by estimating the displacements of the glenoid in each image and correcting the shoulder position accordingly. This elimination allows for a more compact model describing the pose of the arm, hence two parameters modelling the six DoF.

The idea is to find a relationship between the displacements of the GH-joint[1] (denoted Δv and Δh) and the angle between the torso and the upper arm, ϕ. For each image ϕ is estimated based on the position of the hand in the image and Δv and Δh are estimated. In section 2.2.1 we show how to estimate the relationship between ϕ and Δv and how to estimate the relationship between the position of the hand and ϕ.

2.2.1 Relating ϕ and Δv

To understand the relationship between ϕ and Δv the shoulder complex is investigated in more detail. The shoulder complex consists of two bones (the shoulder girdle); the scapula and the clavicle, see figure 2. The former is the large bone on the back of the shoulder and the latter is the one from the breastbone to the top of the shoulder tip (Codman, 1934). The clavicle is a long thin bone connected to the breastbone in the sterno-clavicular joint. It functions as a ball-and-socket joint and is denoted the SC-joint, see figure 2. In the other end the clavicle is connected to the scapula in the acromio-clavicular joint that also functions as a ball-and-socket joint. This joint is denoted the AC-joint. The scapula is a large triangular bone, which contains the glenoid below the AC-joint. The scapula is not

[1] The exact location of the GH-joint is defined as the centre of the humerus head that is the point the humerus rotates about, hence its position is fixed with respect to the glenoid. This means that finding the displacement of the GH-joint is equivalent to finding the displacements of the glenoid. The reason for choosing the GH-joint over the glenoid is that the former has a more clearly anatomic definition.

connected directly to the thorax through a joint but instead via muscles. This results in a very flexible "joint" which can both rotate and translate with respect to the thorax.

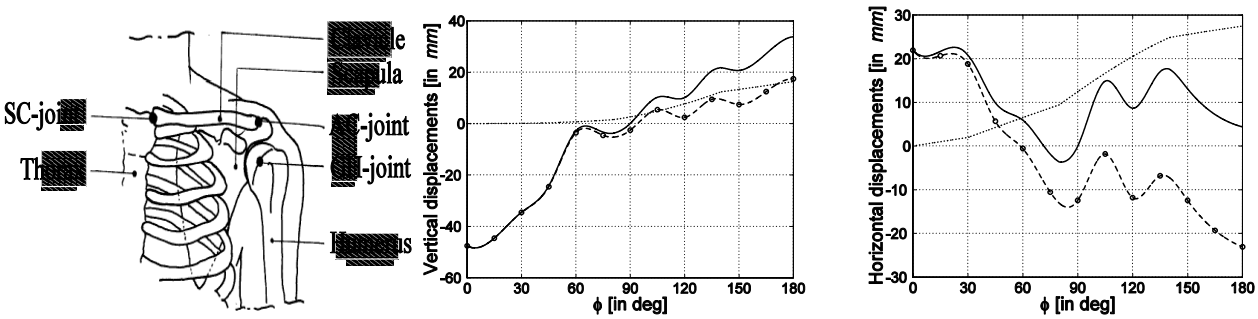


Figure 2. Left: The different bones and joints in the shoulder complex. Figure after (Breteler *et al.*, 1999). Middle and Right: The vertical and horizontal displacements, respectively, as a function of □. The dashed graphs are the displacements of the AC-joint due to the rotation in the SC-joint. The circles are the actual measurements from (Dvir & Berme, 1978) while the graphs are obtained by spline interpolations. The dotted graphs are the additional displacements of the GH-joint due to rotation in the AC-joint. The solid graphs are the total displacements of the GH-joint with respect to the SC-joint normalised to zero at $\phi = 90^{o}$

The value of Δv is the same as the vertical displacement of the GH-joint with respect to the resting position where $\phi = 0^{o}$ and originates from rotation in the SC-joint and the AC-joint. Actually, without the rotation in the SC-joint and AC-joint the elevation of the arm would be limited. When elevating the arm to 180^{o} only 120^{o} comes form the rotation in the GH-joint and the rest comes from the rotation of the scapula (Culham & Peat, 1993), hence in the SC-joint and the AC-joint.

The primary displacement of the GH-joint comes from the elevation of the AC-joint, hence the upward rotation in the SC-joint. In the work by (Dvir & Berme, 1978) measurements describing the displacement of the AC-joint as a function of ϕ□are presented, see the dashed graph in figure 2 (middle). To obtain a complete relationship we need to add Δy that expresses the vertical displacement of the GH-joint originating from the rotation in the AC-joint. We assume that the clavicle is horizontal when $\phi = 0^{o}$ and that the GH-joint is located *r mm* directly below the AC-joint. Δy, can then be calculated as $\Delta y = r\,(1 - \sin(\tau))$ where $\tau = 270^{o} - \alpha - \beta$. α and β are the rotations in the SC-joint and AC-joint, respectively (Moeslund, 2003). *r* is approximately equal to 10.4% of the length of the upper arm (Leva, 1996). That is, $r = A_U \cdot 0.104$.

To be able to calculate Δy as a function of ϕ we need to known how α and β depend on ϕ. The relationship is via the rotation of the scapula, which is therefore investigated further. To structure our investigation we utilise the concept of the "shoulder rhythm" (Zatsiorsky, 1998) or "scapulohumeral rhythm" (Culham & Peat, 1993). The shoulder rhythm is defined as the ratio, *R*, of the upwards rotation in the GH-joint, η, and the angle between the scapula and torso, ψ, hence $R = \eta / \psi$.. Since $\phi = \eta + \psi$ we can apply the ratios reported in the literature to find the relationship between the rotation of the scapula and ϕ as $\psi = \dfrac{\phi}{R+1}$

In average the ratio of an 180^{o} elevation of ϕ is 2:1 (Inman *et al.*, 1944). However, the ratio wary a lot during a 180^{o} elevation. Usually it is divided into four phases (Culham & Peat, 1993). From the ratios in these four phases ϕ can be calculated. Hence, the relationship

between α and ϕ, and β and ϕ, respectively, can be derived and the following values of □ can be calculated (Moeslund, 2003)

$$\tau = \begin{cases} 90.0^o - 0.12\phi & \text{if } \phi \in [0^o, 30^o[\\ 94.8^o - 0.28\phi & \text{if } \phi \in [30^o, 80^o[\\ 119.2^o - 0.59\phi & \text{if } \phi \in [80^o, 140^o[\\ 68.5^o - 0.22\phi & \text{if } \phi \in [140^o, 180^o] \end{cases} \quad (2)$$

In figure 2 (middle) $\Delta y(\phi)$ is shown as the dotted graph. The solid graph is the sum of the dashed- and dotted-graphs, hence $\Delta v(\phi)$. The relationship between ϕ□and □Δh which can be found in a similar manner (Moeslund, 2003) is illustrated in figure 2 (right).

After having related the two prismatic joints to ϕ we need to relate ϕ to the parameters of the local screw axis model. For H_Z this is done by projecting the predicted position of the hand onto the camera ray through the hand. For α this is done using prediction and inverse kinematics (Moeslund, 2003).

3. Kinematic Constraints

Even though the local screw axis model is very compact it still has a large solution-space. Therefore constraints are introduced to prune the solution-space.

When a human moves his/her arm two types of constraints ensure a plausible sequence of poses. These are kinematic constraints and dynamic constraints. The former is concerned with the position and all higher order derivatives of the position variable(s). These constraints are defined without regard to the forces that cause the actual motion of the arm, and are measurable in the image(s). The dynamic constraints on the other hand require knowledge about anatomic features that are not directly measurable, for example masses of bodies and strength of muscles. Furthermore, humans first consider the kinematics and then the dynamics when positioning the arm (Kondo, 1991). This suggests that the pruning effect from the kinematic constraints is dominant.

The principal kinematic constraints come from the limits on the joint angles, e.g., that the arm cannot bend backwards at the elbow, and are defined in anatomy. The actual values of these limits are, however, not universal and differ between individuals. In table 1 the limits for the first author is listed and in figure 1 the angles are illustrated.

	θ_1	θ_2	θ_3	θ_4
Minimum	-135^o	-135^o	0^o	45^o
Maximum	45^o	100^o	145^o	180^o

Table 1. The legal ranges of the different joint angles

Besides angle values also the angle velocity and acceleration yield constraints. Their ranges depend on the activity carried out. Here "normal" movements are considered with a maximum velocity of $400^o/s$. It is assumed that a subject can accelerate the upper- and lower arm to their maximum angle velocity within one tenth of a second, i.e., the maximum acceleration is $4000^o/s^2$.

In the following the limits on the joint angles and their derivatives together with geometric reasoning are applied to prune H_Z and α. The pruning is done through six constraints: four pruning H_Z and two pruning α, see (Moeslund, 2003) for a more detailed description of the constraints.

3.1 Pruning H_Z Using Static Constraints

First H_Z is pruned through three static constraints. The first constraint states that the distance between the 3D hand position and the shoulder needs to be less than the total length of the arm; hence a sphere limits the hand position. To calculate the allowed interval for H_Z we find where the line in equation 1 intersects the sphere, yielding $P_Z + t_2 \cdot D_Z \leq H_Z \leq P_Z + t_1 \cdot D_Z$, where t_1 and t_2 define the intersection points.

The second constraint is an angle constraint. When the position of the hand in the image is to the left of the shoulder ($H_X > 0$) H_Z can be pruned using the limits on θ_1. The limitations on the angle mean, among other things, that the upper arm can only rotate 45^o backwards. Together with the limitations of the other angles the minimum H_Z positions of the hand is limited by θ_1 in the following way: $\tan(45^o) \geq \frac{H_Z}{H_X}$, where the angle is measured from the x-axis. Inserting the parametric equation of the line, equation 1, and isolating *t* allows us to calculate the smallest value of H_Z as $H_{Z,\min} = P_Z + t \cdot D_Z$

The final static constraint pruning H_Z is an occlusion constraint. When the position of the hand in the image is to the right of the shoulder $H_X \leq 0$ the hand is likely to be in front of the head or torso. If this is the case H_Z can be pruned by finding the intersection between the line, *l*, in equation 1 and a representation of the head and torso, respectively. The torso is modelled using an elliptic cylinder with its semi-axes (*a* and *b*) parallel to the *x* and *z*-coordinate axes shown in figure 1. The head is modelled as an ellipsoid with semi-axes *i*, *j*, and *k*, where *i*=*k* (Moeslund, 2003). The parametric line in equation 1 is inserted into both the equation of the ellipsoid and elliptic cylinder and solved with respect to *t*. If *t* is real and its *y*-value belongs to one of the shapes, the limits on H_Z is found as

$$H_{Z,\min} = P_Z + t \cdot D_Z$$

3.2 Pruning H_Z Using a Temporal Constraint

The final pruning of H_Z is based on a temporal constraint stating that the displacement of H_Z between two consecutive frames is bounded by the limits on the joint angles[2].

For each image the limits can be tightened using the previously estimated angle values together with the limits on the velocity and acceleration. Altogether a new set of legal intervals for the four angles is obtained, namely Φ_1, Φ_2, Φ_3, and Φ_4. Each interval is calculated as

$$\Phi = \left[\Phi_{\min}, \Phi_{\max}\right] = \left[\max\left\{\theta_{\min}, \theta^* + \Delta^- \theta\right\}, \min\left\{\theta_{\max}, \theta^* + \Delta^+ \theta\right\}\right] \quad (3)$$

Where θ^* is the value of the joint angle in the previous frame, $\theta_{\min}$ and $\theta_{\max}$ are defined in table 1, and $\Delta^- \theta$ and $\Delta^+ \theta$ are defined in figure 3 (Moeslund, 2003).

In some situations, e.g., when $\theta^* = \theta_{\max} \wedge V^* > 200^O / s$ a sudden stop of the movement of the hand is required. Obviously, this is not realistic and a minimum interval wherein the hand is bound to be is defined to be 20^O. Equation 3 is therefore expanded to

[2] This is calculated for a frame-rate of *10Hz*

$$\Phi = \begin{cases} \left[\theta_{\max} - 20^{o}\,, \theta_{\max}\right] & \text{if } \left(\Phi_{\max} - \Phi_{\min}\right) < 20^{o} \wedge \Phi_{\min} > \theta_{\min} \\ \left[\theta_{\min}\,, \theta_{\min} + 20^{o}\right] & \text{if } \left(\Phi_{\max} - \Phi_{\min}\right) < 20^{o} \wedge \Phi_{\max} > \theta_{\max} \\ \left[\Phi_{\min}\,, \Phi_{\max}\right] & \text{otherwise} \end{cases} \tag{4}$$

The maximum change of H_Z between two frames occurs when the arm moves in the (y=0)-plane and can be found as $H_Z^* - \delta \le H_Z \le H_Z^* + \delta$, where H_Z^* is the value of H_Z in the previous frame and δ is the maximum change of H_Z found by investigating the differential geometry of the arm within the current range of the joint angles Φ_1 and Φ_4 (Moeslund, 2003).

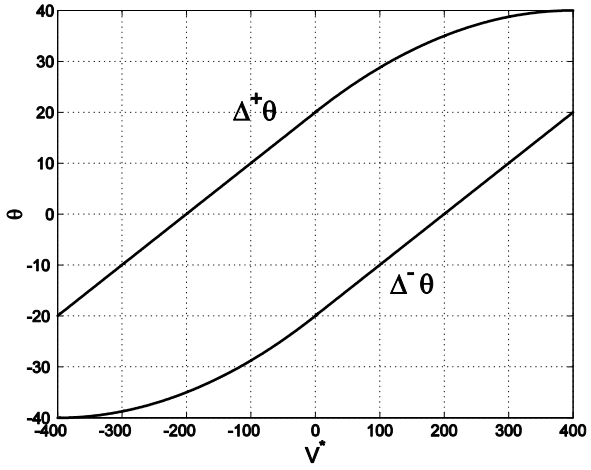


Figure 3. The upper and lower limits on the joint angles as a function of the velocity in the last frame, V^*

3.3 Pruning α Using Joint Angle Constraints

This constraint prunes α by mapping the legal intervals for the joint angles $(\Phi_1, \Phi_2, \Phi_3, \Phi_4)$ to α, Φ_α. To make the following description conceptually easier the arm configuration is viewed as a triangle spanned by the shoulder, elbow, and hand. In terms of angles within the triangle this will make a number of configurations equal. Φ_4 only apply in the interval [0^{o} , 180^{o}] and angles outside this interval will be constrained by Φ_3. Independently of α and the three other angles, $\theta 4$ needs to be within Φ_4. If not, the current value of H_Z can be ignored altogether, i.e., the entire range of α is pruned. The current value of $\theta 4$ is calculated using the cosine-relation (Moeslund, 2003).

A relationship between the remaining three angles and α is found by defining a reference triangle, where $\vec{E} = (A_U, 0, 0)$ and $\vec{H} = (A_U + A_L \cdot \sin(\theta_4), -A_L \cdot \cos(\theta_4), 0)$, and explaining how to rotate it into the current triangle spanned by the shoulder, the elbow, and the hand. The rotation is described in two ways. One is by using the three joint angles R_θ and another using $R(\alpha)$. The two rotations will be equal yielding the relationship between the joint angles and α. The three angles, see figure 1, are Y-Z-X Euler angles, hence

$$R_\theta = R(\theta_1, \theta_2, \theta_3) = \begin{bmatrix} c1 \cdot c2 & -c1 \cdot s2 \cdot c3 + s1 \cdot s3 & c1 \cdot s2 \cdot s3 + s1 \cdot c3 \\ s2 & c2 \cdot c3 & -c2 \cdot s3 \\ -s1 \cdot c2 & s1 \cdot s2 \cdot c3 + c1 \cdot s3 & -s1 \cdot s2 \cdot s3 + c1 \cdot c3 \end{bmatrix} \tag{5}$$

where $c1=\cos(\theta_1)$ and $s2=\sin(\theta_2)$ etc. The other rotation, R(α), is obtained by performing a number of rotations. First, the shoulder-hand-line of the reference triangle is rotated to be aligned with the x-axis. Next, this line is rotated to be aligned with the current shoulder-hand-line, $\vec{H}$. This is done by first aligning the z-component and then the y-component. The shoulder-hand-line of the reference triangle is now aligned with the shoulder-hand-line of the current configuration. The final rotation is to align the position of the elbow in the reference triangle with the current elbow position. This is done using Rodriques' formula (Craig, 1989), which rotates the reference triangle around, $\vec{H}$ by α degrees. The resulting rotation matrix is quite complicated, but each entry can be expressed as: $a \cdot \sin(\alpha) + b \cdot \cos(\alpha) + c$ where *a*, *b*, and *c* are constants.

We can now use the fact that $R(\theta_1, \theta_2, \theta_3) = R(a)$ to calculate how the limits on the Euler angles can prune α. First we see how θ_2 prune α. We apply entry (2,1) yielding: $\sin(\theta_2) = a_{21} \cdot \sin(\alpha) + b_{21} \cdot \cos(\alpha) + c_{21}$

Looking at this equation as a function of α we can see that the right-hand side is a sine curve and the left-hand side is two straight lines; corresponding to the minimum and maximum allowed values of θ_2 defined by Φ_2. The equation can thou yield six different results, i.e., six different types of pruning intervals. These are illustrated in figure 4, where the shaded area (from zero to 360) illustrates the legal values for θ_2, and the shaded region(s) on the α-axis illustrates the legal α-values. For details see (Moeslund, 2003).

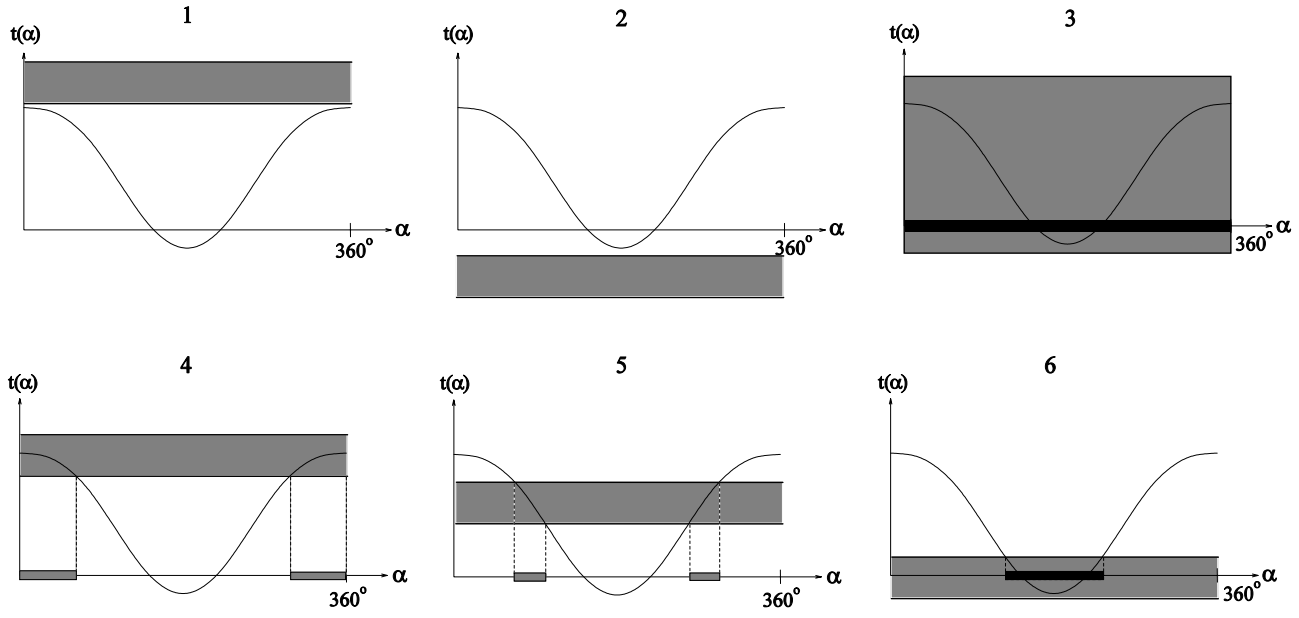


Figure 4. The six possible situations when combining the two rotation matrices

To calculate how θ_1 prunes α and how θ_3 prunes α entry (1,1) and (2,2) are applied, respectively. Since two joint angles are present in these equations, the joined legal intervals for each equation are defined over two variables rather than one, hence a legal region. This can result in more legal intervals for α due to more complex monotonic characteristics, but otherwise the calculations are similar to those relating θ_2 and α For details see (Moeslund, 2003).

3.4 Pruning α Using a Collision Constraint

The fact that two body parts cannot occupy the same space at the same time is the final constraint that prunes α. This constraint is applied for each non-pruned H_Z-value and it is therefore important to have an analytic solution that evaluates all α-values at a time instead of one at the time. To obtain this two simplifications are introduced. First, a collision is defined as a configuration where the elbow is colliding with the torso or head. Second, the torso and head are modelled as one square having sides equal to those used to define the torso in section 3.1, except $-\infty < y < \infty$. The impact of these simplifications is in practice minor and therefore justifiable.
Before applying the model two heuristic rules are introduced. If the current pose contains a collision then

- the distance in the *z*-direction between the hand and torso needs to be less than the length of the lower arm: $H_Z - b < A_l$, or
- the distance in the *x*-direction between the hand and torso needs to be less than the length of the lower arm: $H_X < A_l$

In other words, one of the two rules has to be fulfilled in order to evaluate the collision constraint. If not, the entire α-circle related to this H_Z-value is pruned.
For the remaining H_Z-values the following is done. For each value of H_Z the elbow describes a circle in space and, in general, an ellipse in the x-z-variables. The pruned interval of α is found as $[\alpha_1, \alpha_2]$ where α_1 and α_2 are the intersection points (if any exist) between the ellipse and the rectangle (ignoring the y-component).

4. Pose Estimation

So far we have focused on deriving a compact representation of the arm and shoulder and pruning this representation with respect to impossible configurations. The result is a highly reduced solution-space for each frame. Now we want to apply the result to an actual computer vision system those task is to capture the pose of the arm. That is, given the highly reduced solution-space, how do we find *the* correct pose of the arm? A brute force solution can some times be applied depending on the resolution in the solution-space and the real-time requirements. But in general and in the case of estimating the pose of the entire human the overall solution-space will still be too large even though the solution-space for the arm is limited, i.e., a non-brute force approach is required. Furthermore, the representation derived above is based on the fact that the human hand can always be found in the correct position in the image. Clearly this will not always be the case and one should therefore expect a certain amount of uncertainty and that the hand sometimes will disappear altogether. A solution to these problems is to imbed the tracking algorithm into the probabilistic framework known as Sequential Monte Carlo (SMC). The SMC framework is an efficient solution to the brute force problem and inherently handles uncertainty. We phrase our pose estimating problem as a matter of including auxiliary information (the position of the hand in the image) into the SMC framework. In the rest of this section we show how the SMC operates and how it can be improved by adding auxiliary information. Concretely, we first present the SMC framework. We then show how to include the auxiliary information. Finally we present a method for comparing the arm model with the image data.

4.1 The SMC Framework

The SMC algorithm is defined in terms of Bayes' rule and by using the first order Markov assumption. That is, the posterior probability density function (PDF) is equal to the observation PDF multiplied by the prior PDF, where the prior PDF is the predicted posterior PDF from time *t-1*:

$$p(\vec{x}_t \mid \vec{y}_t) \propto p(\vec{y}_t \mid \vec{x}_t)\, p(\vec{x}_t \mid \vec{y}_{t-1}) \tag{6}$$

where $\vec{x}_t$ is the state and $\vec{y}_t$ contains the image measurements. The predicted posterior PDF is defined as

$$p(\vec{x}_t \mid \vec{y}_{t-1}) = \int p(\vec{x}_t \mid \vec{x}_{t-1})\, p(\vec{x}_{t-1} \mid \vec{y}_{t-1}) d\vec{x}_{t-1} \tag{7}$$

where $p(\vec{x}_t \mid \vec{x}_{t-1})$ is the motion model governing the dynamics of the tracking process, i.e., the prediction, and $p(\vec{x}_{t-1} \mid \vec{y}_{t-1})$ is the posterior PDF from the previous frame. The SMC algorithm estimates $p(\vec{x}_t \mid \vec{y}_t)$ by selecting a number, *N*, of (hopefully) representative states (particles) from $p(\vec{x}_{t-1} \mid \vec{y}_{t-1})$, predicting these using $p(\vec{x}_t \mid \vec{x}_{t-1})$, and finally giving each particle a weight in accordance with the observation PDF. In this work the state vector, $\vec{x}_t$, represents the 3D model of the arm, i.e., (α, H_z).

The observation PDF, $p(\vec{y}_t \mid \vec{x}_t)$, expresses how alike each state and the image measurements are. In this work the image measurements are the probabilities of the orientations of the upper and lower arm in the image, respectively, i.e., $\vec{y}_t = [p_u(y_u), p_l(y_l)]^T$, where $p_u(y_u)$ and $p_l(y_l)$ are the PDFs of the different orientations of the upper and lower arm in the image, respectively. We define the observation PDF as

$$p(\vec{y}_t \mid \vec{x}_t) = \frac{p_u(y_u(\vec{x}_t)) + p_l(y_l(\vec{x}_t))}{2} \tag{8}$$

where $y_u(\vec{x}_t)$ and $y_l(\vec{x}_t)$ map from (α, H_Z) to the orientation of the upper and lower arm in the image, respectively[3].

4.2 Including the Auxiliary Information

In this section we describe how including auxiliary information enhances the SMC algorithm. The auxiliary information is in the form of the position of the hand in the image. Firstly we will describe how the auxiliary information is obtained and related to the SMC algorithm. Secondly we will describe how to apply the auxiliary information to correct the states of the predicted particles.

4.2.1 Obtaining the Auxiliary Information

The hand candidates in an image are detected based on skin colour segmentation. We first convert each image pixel, (R,G,B), into chromaticity, (r,g,b), and make a binary image

[3] These mappings require the camera parameters as well. But to enhance the concept we have left them out of the expression.

based on a Mahalanobis classifier, those mean and covariance are found during training. In the binary image we apply morphology followed by a connected component analysis. This gives us a number, m, of skin blobs, b_i, which each could represent the hand. Each skin blob is used to correct a number of particles according to the likelihood of this particular blob being a hand, $p(hand \mid b_i)$. That is, the number of particles, N, available at each time instance are divided between the m skin blobs so that blob b_i is associated with p_i particles, where

$$p_i = N \frac{p(hand \mid b_i)}{\sum_{i=1}^{m} p(hand \mid b_i)} \tag{9}$$

The problem with this approach is that it assumes that the correct hand position always is among the detected skin blobs. When this is not the case the entire system is likely to fail. To overcome this, we adapt the approach taken in (Davis *et al.*, 2000), where only a portion of the N particles are predicted and the remaining particles are drawn from a uniform distribution. Similar, we will always let $T \cdot N$ particles be predicted and weighted regardless of the auxiliary information, i.e., N is replaced by $N - T \cdot N$. Concretely we say that at least 10% of the particles should be drawn without regard to the auxiliary information and define T as

$$T = \begin{cases} 0.1, & \text{if } k > 0.9 \\ 1 - k, & \text{else} \end{cases} \tag{10}$$

where k is the likelihood of the skin blob most likely to represent the hand, i.e., $k = \arg\max_i \{p(hand \mid b_i)\}$

4.2.1.1 Defining the Likelihood of a Hand

We define the likelihood of the hand as

$$p(hand \mid b_i) = F\left(\prod_{j=1}^{t} w_j \cdot p(hand \mid f_j, b_i) \right) \tag{11}$$

where $F(\cdot)$ scales the likelihood, t is the number of features, w_j is the weight of the jth feature, f_j is the jth feature, and $p(hand \mid f_j, b_i)$ is the likelihood of the hand given the jth feature and the ith skin blob. The scaling of the likelihood is necessary as we use this value not only as a relative measure, but also as an absolute measure when defining T. In this work $F(x) = 1 - exp(-5x)$ was found to work well. We use three equally weighted features, i.e., $t=3$ and $w_j=1$. The first feature is based on the number of pixels in the blob. As this feature is dependent on a number of different aspects, such as the distance to the camera, we apply this feature in a very conservative manner

$$p(hand \mid f_1, b_i) = \begin{cases} 0, & \text{if } A > TH_{\max} \\ 0, & \text{if } A < TH_{\min} \\ 1, & \text{else} \end{cases} \tag{12}$$

where TH_{min} and TH_{max} define the lower and upper limits, respectively, and A is the area, i.e., the number of pixels.
The second feature is based on the idea that the centre of gravity (CoG) and the centre of the hand should be close to each other. This is evaluated by estimating the centre of the blob (hand) by a distance transform and comparing it with the CoG in the following way

$$p(hand \mid f_2, b_i) = 1 - \left(\frac{DT_{\max} - d(CoG)}{DT_{\max}} \right) \tag{13}$$

where *d(CoG)* is the value found by the distance transform in the position of the CoG and DT_{max} is the maximum value found by the distance transform inside the blob. The last feature is inspired by the fact that an ellipse often can model the shape of a hand. We therefore calculate the semi-axes of the ellipse that corresponds to the area and perimeter of the blob. This ellipse is denoted E_d and compared to the blob to see how well it matches. An ellipse can be described by its area $A = ab\pi$ and perimeter $P = 2\pi\sqrt{\frac{1}{2}\left(a^2 + b^2\right)}$, where $2a$ is the major axis and $2b$ is the minor axis. Expressing the major and minor axes in terms of A and P yields

$$\begin{aligned} 2a &= \sqrt{\frac{P^2}{2\pi^2} + \frac{2A}{\pi}} + \sqrt{\frac{P^2}{2\pi^2} - \frac{2A}{\pi}} \\ 2b &= \sqrt{\frac{P^2}{2\pi^2} + \frac{2A}{\pi}} - \sqrt{\frac{P^2}{2\pi^2} - \frac{2A}{\pi}} \end{aligned} \tag{14}$$

The measured area and perimeter of the blob are used to calculate the axes of E_d, a and b. The centre of E_d is then placed in the CoG and E_d is rotated and compared with the blob. The rotation is done in a coarser-to-finer manner and the comparison is carried out by calculating the intersection divided by the union of the two regions, that is

$$p(hand \mid f_3, b_i) = \arg\max_{\delta} \left\{ \frac{E_d(\delta) \cap A}{E_d(\delta) + A - E_d(\delta) \cap A} \right\} \tag{15}$$

where δ is the rotation of the ellipse E_d, and A is the area of the blob.

4.2.2 Applying the Auxiliary Information

In this subsection we describe how one particle is corrected based on the auxiliary information. We first convert (α,H_Z) into the 3D position of the elbow, $\vec{E}$, and hand, $\vec{H}$, respectively. We do this conversion for two reasons. Firstly, more smooth trajectories can be expected for these parameters and hence, better motion models can be defined. Secondly, we can directly apply the CoG of the hand to correct the predictions which is not so easy in the (α,H_Z) representation. After this conversion both $\vec{E}$ and $\vec{H}$ are predicted using a linear first order motion model and then kinematic constraints described earlier are applied to ensure a possible configuration.
First we will show how the prediction of the hand, $\vec{H}$, is corrected and hereafter we will show how the predicted position of the elbow, $\vec{E}$, is corrected. In figure 5 the predictions are illustrated using subscript 'p' while the corrected predictions are illustrated using

subscript 'c'. As mentioned earlier we assume a calibrated camera and can thou span a line in 3D, l , via the CoG and the camera, see figure 5 and equation 1. We can therefore correct the prediction by projecting the predicted position of the hand, $\vec{H}_P$, to the line, l . The projected prediction is denoted $\vec{H}_1$ and calculated $\vec{H}_1 = \vec{P} + \left((\vec{H}_P - \vec{P}) \cdot \vec{D}\right)\vec{D}$ where $\vec{P}$ and $\vec{D}$ are the line parameters of *l*, see equation 1.

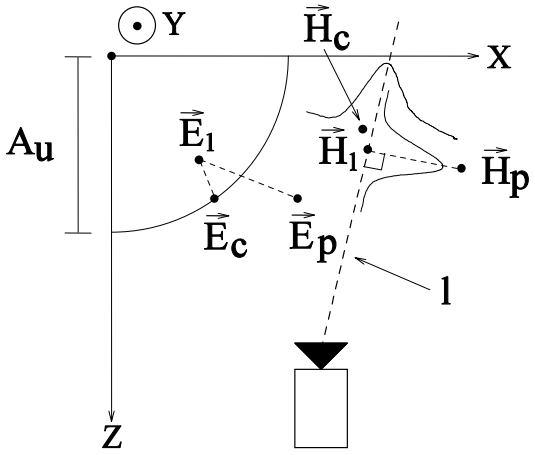


Figure 5. The shoulder coordinate system seen from above. The circle arc illustrates the sphere that defines the possible positions of the elbow. The large dashed line illustrates a camera ray through the hand. See the text for a definition of the parameters

We randomly diffuse $\vec{H}_1$ to account for the uncertainties in the estimate of the CoG. Concretely we first randomly draw a value from a Gaussian distribution with mean in $\vec{H}_1$ and standard deviation as $c \cdot \left\|\vec{H}_P - \vec{H}_1\right\|$, where c is a predefined constant. This point defines the displacement along the camera ray with respect to $\vec{H}_1$. The resulting point is now rotated around the vector $\overrightarrow{H_1 H_P}$ using Rodriques' formula (Craig, 1989). The number of rotated degrees is determined by randomly sampling from a uniform distribution. The final diffusion of the point is along the vector $\overrightarrow{H_1 H_P}$. The displacement is define with respect to $\vec{H}_1$ and designed so that the maximum probability is at $\vec{H}_1$ and that the tail towards $\vec{H}_P$ is more likely than on the opposite side. This corresponds to a Poisson distribution with its maximum probability located in $\vec{H}_1$. We implement this by combining two Gaussian distributions, G_1 and G_2, each with mean in $\vec{H}_1$. G_1 represents the distribution on the opposite side of $\vec{H}_P$ and its variance is controlled by $p(hand \mid b_i)$. G_2 represents the distribution on the same side as $\vec{H}_P$ and its variance is controlled by both $\left\|\vec{H}_P - \vec{H}_1\right\|$ and $p(hand \mid b_i)$. In praxis we first choose from which side of the mean we should draw a sample and then draws it from the appropriate (one-sided) Gaussian distribution. After these three diffusions we have the final corrected particle, denoted $\vec{H}_c$. The difference between the predicted and corrected particles yields a measure of the prediction error: $\vec{H}_e = \vec{H}_c - \vec{H}_P$.

The predicted position of the elbow cannot directly be corrected by the auxiliary information. However, we know the elbow is likely to have a prediction error closely related to that of the hand, as the hand and elbow are part of the same open-looped kinematic chain. We therefore calculate the corrected position, $\vec{E}_c$, by first adding the

prediction error of the hand to the predicted value of the elbow, yielding $\vec{E}_1 = \vec{E}_P + \vec{H}_e$, and then finding the point closest to $\vec{E}_1$ that results in a legal configuration of the arm. In mathematical terms $\vec{E}_c = \arg\min_{\vec{E}} \left\| \vec{E} - \vec{E}_1 \right\|$ subjected to the constraints $\left\| \vec{E} \right\| = A_U$ and $\left\| \overrightarrow{EH_c} \right\| = A_L$. The solution to this problem can be found in (Moeslund, 2003).

4.3 The Observation PDF

For each blob, b_i, we estimate an observation PDF and use this to weight the particles related to a particular blob. For the $T \cdot N$ particles that are not related to a blob we use the less accurate approach of chamfer matching instead of equation 8. The distance transform is calculated on the edge image in figure 6.

The observation PDF in equation 8 is based on the orientations of the arm segments in the image. We estimate the PDFs of the orientations of the upper arm, $p_u(y_u)$, and lower arm, $p_l(y_l)$, respectively, based on edge pixels. As our input images contain background clutter and non-trivial clothes we utilize temporal edge pixels[4]. That is, we find the edge pixels in the current image using a standard edge detector and AND this result with the difference image achieved by subtracting the current- and the previous image, see figure 6 for an example[5]. Those pixels actually belonging to the arm will be located in four classes, two for the upper arm and two for the lower arm, respectively.

Our system does not impose restrictions on the clothes of the user. The clothes will in general follow gravity, hence the two classes of pixels originating from the upper sides (with respect to gravity) of the upper- and lower arm will model the structure of the arm better, see figure 6. We therefore only consider temporal edge pixels located on the ”upper” sides. Concretely we define ”upper” and ”lower” via two lines described by the position of the shoulder in the image, the CoG of the blob, and the predicted position of the most plausible elbow location found among the most likely in the previous frame.

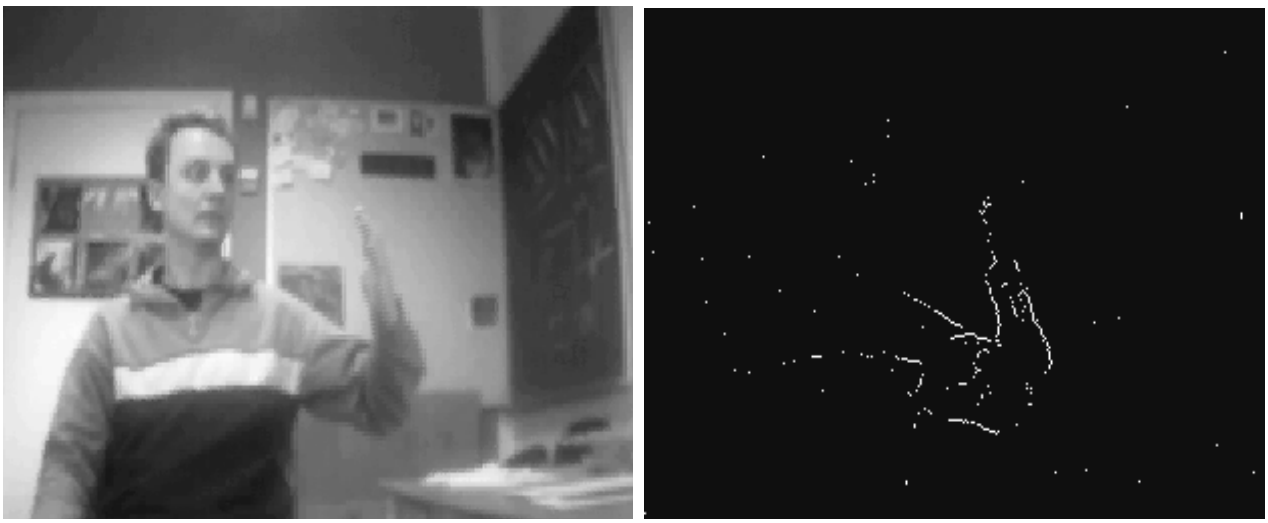


Figure 6. Left: A typical input image Right: The temporal edge pixels

[4] We assume that only ego-motion is present.

[5] When only a few temporal edges can be found we conclude that on motion is present and do not update the state-space parameters, i.e., no processing beyond this point is carried out.

We wish to estimate $p_u(y_u)$ and $p_l(y_l)$ independently and therefore separate the temporal edge pixels into two groups by calculating the perpendicular distance from each pixel to the two lines. As the prediction of the position of the elbow is uncertain we ignore all pixels within a certain distance from the predicted position of the elbow. Furthermore, we ignore all pixels too far away from both lines. When no predictions are available we use chamfer matching (as described above) as the observation PDF.

Estimating the orientation of a straight line from data can be carried out in different ways, e.g., via principal component analysis or linear regression. However, as we will not model the distribution of the orientations via Gaussians we cannot apply these methods. Instead we apply the dynamic Hough transform (DHT). It estimates the likelihood of each possible orientation, hence allowing multiple peaks in the observation PDF. The choice of the DHT is furthermore motivated by the fact that it adapts to the data. The DHT randomly samples two pixels from one group and calculates the orientation of the line spanned by the two pixels. The more times the groups are sampled the better the estimation of the PDFs will be. On the other hand many samplings also lead to large processing time. The sampling of one group is therefore terminated as soon as the variance of the PDF is stable. To evaluate the stability of the variance after *n* samplings the variance of the last *j* variances is calculated as

$$\nu_{jn}^2 = \frac{1}{j}\sum_{i=n-j}^{n}\left(\sigma_i^2 - \mu_{jn}\right)^2 \tag{16}$$

where σ_i^2 is the variance after *i* samplings and μ_{jn} is the mean of the last *j* variances.

The stop criterion is defined as the number of samplings, *n*, where the last *j* samplings are within the interval $\left]\mu_{jn} - \lambda, \mu_{jn} + \lambda\right[$. The distribution of the last *j* variances will in general follow a Uniform distribution. The theoretical variance of such a distribution in the given interval can be estimated as $\lambda^2/12$ (Ross, 1987). When the mean of the variances, μ_{jn} is large it indicates large uncertainty in the PDF, which again indicates weak lines in the temporal edge image. A stable variance for such a PDF tends to require a larger value of λ compared to an image with stronger lines. To account for this difference λ is defined with respect to μ_{jn} as

$$\lambda = \frac{\mu_{jn}}{\gamma} \tag{17}$$

where γ is found empirically. Setting the estimated variance equal to the theoretical variance yields $\lambda = \nu_{jn}\sqrt{12}$. Inserting this result into equation 17 and writing it as an inequality yields

$$\nu_{jn}^2 \leq \frac{\mu_{jn}^2}{12 \cdot \gamma^2} \tag{18}$$

Altogether the stop criterion is found as the smallest *n* for which inequality 18 is true. To speed up the calculations the variance is not recalculated after each new sampling, but rather for every *10th* sampling. Using the above-described procedure we obtain two independent PDFs, one for the upper arm, $p_u(y_u)$ and one for the lower arm, $p_l(y_l)$. Different number of samplings might have been used to estimate the two PDFs. The accumulated probability mass for each PDF is therefore normalized to 1.

5. Results and Discussion

In this section we will evaluate our approach. That is, we will present results and discuss: the local screw axis model, the effect of the constraints, and the improved SMC framework.

5.1 The Arm Model

Modelling the pose of the arm, i.e., the GH-joint and elbow joints, by α and H_Z is a novel approach. Through geometric reasoning it can easily be shown that there exists a one-to-one mapping between our representation and the standard representation via four Euler angles. Since the Euler angles representation is sound, the same must be true for the local screw axis model.

Comparing the size of its solution-space with that of the standard approach namely the four Euler angles carries out a quantitative evaluation of the local screw axis model. In table 2 the sizes of the two representations are listed for different resolutions. The calculations are done for standard arm lengths, $A_u = A_l = 30cm$, and no constraints whatsoever are used, thus yielding the full size solution-space. The Greek letters τ and ρ represent the resolution of the Cartesian coordinates and angle values, measured in cm^{-1} and $degrees^{-1}$, respectively.

Model name	Parameters	Size of solution-space	$\tau = \rho = 10$	$\tau = \rho = 1$	$\tau = \rho = 0.1$
Euler angles	$\theta_1, \theta_2, \theta_3, \theta_4$	$(\tau \cdot 360)^4$	$1.68 \cdot 10^{14}$	$1.68 \cdot 10^{10}$	$1.68 \cdot 10^{6}$
Local screw axis model	α, H_Z	$\tau \cdot 360 \cdot 2(\rho A_U + \rho A_l)$	$4.32 \cdot 10^{6}$	$4.32 \cdot 10^{4}$	$4.32 \cdot 10^{2}$

Table 2. Comparing the local screw axis model with the four Euler angles

From the table it can be seen that a huge reduction in the total number of different arm configurations is achieved. In fact, the reduction factors for the three resolutions are: $3.89 \cdot 10^{7}$, $3.89 \cdot 10^{5}$, and $3.89 \cdot 10^{3}$, respectively.

5.2 Effects of the Constraints

How much a particular constraint prunes the solution-space depends on the current position of the hand and the previous estimated position of the arm, in other words spatial and temporal information[6]. It is therefore not possible to state a general pruning effect but in table 3 the intervals of the pruning effects are shown together with the average effects (Moeslund, 2003).

The first four constraints usually overlap, except for the second and third which are mutually exclusive. They, however, each overlap with the two others. The last two constraints might also overlap each other and it is therefore not possible to calculate a general accumulated effect of the different constraints. Instead the minimum, maximum,

[6] The effect of the three first and the last constraints also depends on the position of the camera, but for simplicity it is assumed that the camera is perpendicular to the torso and infinitively far away.

and average effect can be estimated. For α□ these are 75%, 100%, and 85%, respectively, and for H_Z these are 49%, 100%, and 80%, respectively. Altogether this yields a minimum pruning effect of 87%, a maximum pruning effect of 100%, and an average pruning effect of 97% (Moeslund, 2003).

Parameter	Type of constraint	Minimum	Maximum	Average
H_Z	Distance	0%	100%	48%
H_Z	Angle	0%	50%	25%
H_Z	Occlusion	0%	57%	10%
H_Z	Temporal	51%	92%	77%
□	Joint angle	75%	100%	88%
□	Collision	0%	100%	30%

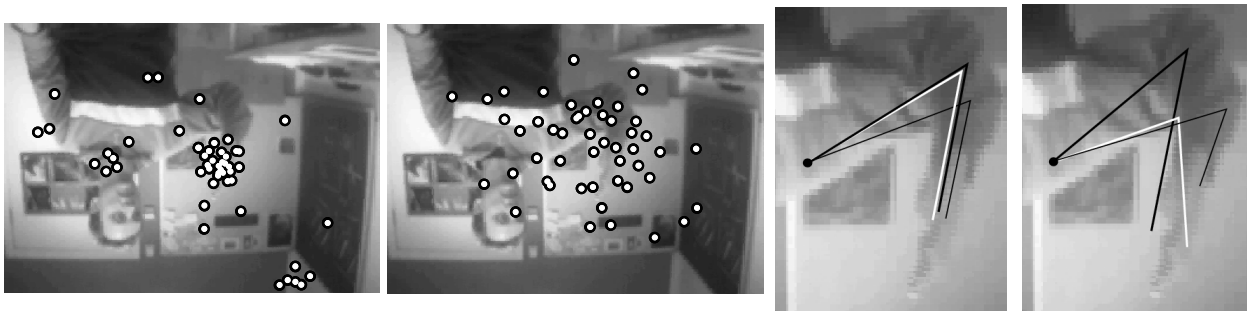
Table 3. The different constraints and their pruning effects.

A resolution of for example *2cm* for H_Z and 5^o for α results in 4320 distinct configurations. Pruning yields in worst-case 553 non-pruned values and in average 125.

5.3 The Improved SMC Framework

The reason for applying the SMC framework is, as described earlier, that the SMC framework can handle both the uncertainties regarding the segmentation of the hand and at the same time avoid the need for an exhaustive search.
Distributing the particles in accordance with the likelihood of the different skin-colour blobs being a hand is in theory a solid approach. It also turns out to be an applicable solution when implementing a SMC-based tracker. In 7.A an example image from a test sequence is shown where 50 particles are used to track the arm. The circles represent the corrected position of the hand projected into the image and it is evident that the main parts of the particles are located on and around the true hand of the hand[7].
As shown above the SMC framework can improve our modelling approach, but in fact our modelling approach can also improve the SMC framework. To illustrate this point we implemented a standard SMC tracker based on the same observation PDF but using Euler angles to represent the solution-space as opposed to the local screw axis model. For both the standard SMC-tracker and our version 50 particles are applied. After tracking the arm for 100 frames the characteristics of the two algorithms are illustrated in figure 7.
In figure 7 we show the values of the predicted particles in the standard SMC algorithm (figure 7.B) and the values of the corrected particles when our algorithm is applied (figure 7.A). We do not visualize the parameters in the solution-space but rather the 3D position of the hand projected into the image. In figure 7.A the main parts of the particles are located around the segmented skin-coloured blobs and especially around the hand. These more focused particles result in a higher probability of finding the correct pose of the arm - even when using as few as 50 particles. This can also be seen in figure 7.C and 7.D where the three particles with the biggest likelihood are illustrated for the standard SMC algorithm (7.D) and when applying our method (7.C). It can be seen that our method improves the results.

[7] Note that the face blob is eliminated by feature, f_1. The neck region is segmented into a different blob and therefore associated with some particles.



A B C D

Figure 7. Next we test whether the true state of the arm is among the states found by the SMC algorithm. If this is the case the SMC algorithm can indeed avoid the need for an exhaustive search. In figure 7.C we show the 2D projection of the three particles with the biggest likelihood (highest weight first: black, white, thin black)

In images such as the one in figure 7.A the posterior PDF is in general ambiguous and a "ridge" will be present in the posterior PDF. This means that a number of correct poses, i.e., poses that can explain the current image data, can be found by increasing the distance between the hand and camera, i.e., moving along the ridge. This tendency can be seen in figure 7.D while the standard SMC algorithm fails to capture this tendency.

6. Conclusion

In this article we have shown how to pose estimate the human arm using monocular computer vision. The hypothesis behind our approach is that the fewer possible solutions and the better the search through these, the higher the likelihood of finding the correct pose in a particular frame. To achieve fewer solutions we did two things. Firstly, we introduced a very compact representation of the human arm based on the position of the hand measured in the current image. We denoted this representation the *local screw axis model*. Secondly, we applied the kinematics of the human arm in order to prune the solution space. In average our constraints can prune the solution-space with 97%

Our representation of the arm is based on the position of the hand in the image. In order to account for the inherent uncertainties in such a representation we imbed our approach in the SMC framework. This framework allows us to model the uncertainties of the position of the hand and the disappearing of the hand in the image (tracking failure).

Besides the above-mentioned issues we have also made a contribution in this work by showing how to model the complex movements in the shoulder without introducing additional parameters. That is, the displacements of the shoulder during arm movement can now be modelled without increasing the dimensionality of the solution-space, i.e., a more precise solution can be achieved. Lastly it should be mentioned that the idea of correcting the predictions in the SMC framework using auxiliary information can in general improved all pose estimating systems where the object to be tracked is an open-kinematic chain.

7. References

An, K.N. & Morrey, B.F. (1985). Biomechanics of the Elbow. *The Elbow and its Disorders*. W.B. Saunders Company, 1985.

Azoz, Y.; Devi, L. & Sharme, R. (1998). Reliable Tracking of Human Arm Dynamics by Multiple Cue Integration and Constraint Fusion. *International Conference on computer Vision and Pattern*. Santa Barbara, California, June, 1998.

Ben-Arie, J.; Wang, Z., Pandit, P. & Rajaram, S. (2002). Human Activity Recognition Using Multidimensional Indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 24, Nr. 7, 2002.

Bregler, C. & Malik, J. (1998). Tracking People with Twists and Exponential Maps. *International Conference on Computer Vision and Pattern Recognition*. Santa Barbara, California, June, 1998.

Breteler, K.; Spoor, C.W. & Van der Helm, F.C.T. (1999). Measuring Muscle and Joint Geometry Parameters of a Shoulder for Modelling Purposes. *Journal of Biomechanics*. Vol. 32, nr. 11, 1999.

Codman, E.A. (1934). *The Shoulder*. Boston: Thomas Todd Co. 1934.

Craig, J.J. (1989). *Introduction to Robotics. Mechanics and Control*. Addison Wesley. 1989.

Culham, E. & Peat, M. (1993). Functional Anatomy of the Shoulder Complex. *Journal of Orthopaedic and Sports Physical Therapy*. Vol. 18, nr. 1, 1993.

Davis, L.; Philomin, V. & Duraiswami, R. (2000). Tracking Humans from a Moving Platform. *International Conference on Pattern Recognition*. Barcelona, Spain, Sept., 2000.

Delamarre, Q. & Faugeras, O. (2001). 3D Articulated Models and Multi-view Tracking with Physical Forces. *Computer Vision and Image Understanding*. Vol. 81, Nr. 3, 2001.

Deutscher, J.; Blake, A. & Reid, I. (2000). Articulated Body Motion Capture by Annealed Particle Filtering. *Computer Vision and Pattern Recognition*. South Carolina, June, 2000.

Dvir, Z. & Berme, N. (1978). The Shoulder Complex in Elevation of the Arm: A Mechanism Approach. *Journal of Biomechanics*. Vol. 11, 1978.

Engin, A.E. & Tumer, S.T. (1989). Three-Dimensional Kinematic Modelling of the Human Shoulder Complex - Part 1: Physical Model and Determination of Joint Sinus Cones. *Journal of Biomechanical Engineering*. Vol. 111, 1989.

Fua, P.; Gruen, A., Plankers, R., D'Apuzzo, N. & Thalmann, D. (1998). Human Body Modeling and Motion Analysis From Video Sequences. *International Symposium on Real Time Imaging and Dynamic Analysis*. Hakodate, Japan, June, 1998.

Gavrila, D.M. & Davis, L.S. (1996). 3-D Model-Based Tracking of Humans in Action: A Multi-View Approach. *Conference on Computer Vision and Pattern Recognition*. San Francisco, USA, 1996.

Goncalves, L.; Bernardo, E.D., Ursella, E. & Perona, P. (1995). Monocular Tracking of the Human Arm in 3D. *International Conference on Computer Vision*. Cambridge, Massachusetts, 1995.

Hogfors, C.; Peterson, B., Sigholm, G. & Herberts, P. (1991). Biomechanical Model of the Human Shoulder Joint - 2. The Shoulder Rhythm. *Journal on Biomechanics*. Vol. 24, nr.

Howe, N.R; Leventon, M.E. & Freeman, W.T. (2000). Bayesian Reconstruction of 3D Human Motion from Single-Camera Video. Advances in Neural Information Processing Systems 12. MIT Press, 2000.

Inman, V.T.; Saunders, J.B. & Abbott, L.C. (1944). Observations on the Function of the Shoulder Joint. *Journal on Bone and Joint Surgery*. Vol. 26, 1944.

Isard, M. & Blake, A. (1998). CONDENSATION - conditional density propagation for visual tracking. *International Journal on Computer Vision*. Vol. 29, nr. 1, 1998.

Kondo, K. (1991). Inverse Kinematics of a Human Arm. *Journal on Robotic Systems*. Vol. 8,

Lee, M.W. & Cohen, I. (2004). Human Upper Body Pose Estimation in Static Images. *European Conference on Computer Vision*. Prague, May 2004.

Leva, P. (1996). Joint Center Longitudinal Positions Computed from a Selected Subset of Chandler's Data. *Journal on Biomechanics*. Vol. 29, 1996.

Maurel, W. (1998). *3D Modeling of the Human Upper Limb including the Biomechanics of Joints, Muscles and Soft Tissues*. Ph.D. Thesis, Laboratoire d'Infographie - Ecole Polytechnique Federale de Lausanne, 1998.

Mitchelson, J. & Hilton, A. (2003). From Visual Tracking to Animation using Hierarchical Sampling. *Conference on Model-based Imaging, Rendering, image Analysis and Graphical special Effects.*, France, March, 2003.

Morrey, B.F. (1985). Anatomy of the Ellow Joint. *The Elbow and its Disorders*. W.B. Saunders Company, 1985.

Moeslund, T.B. & Granum, E. (2001). A Survey of Computer Vision-Based Human Motion Capture. *Journal on Computer Vision and Image Understanding*, Vol. 81, nr. 3, 2001.

Moeslund, T.B. (2003). *Computer Vision-Based Motion Capture of Body Language*. Ph.D. Thesis, Laboratory of Computer Vision and Media Technology, Aalborg University, Denmark, 2003.

Ogaki, K.; Iwai, Y. & Yachida, M. (2001). Posture Estimation Based on Motion and Structure Models. *Systems and Computers in Japan*, Vol 32, Nr. 4, 2001.

Ong, E.J. & Gong, S. (1999). Tracking Hybrid 2D-3D Human Models from Multiple Views. *International Workshop on Modeling People at ICCV'99*. Corfu, Greece, September, 1999.

Pavlovic, V.; Rehg, J.M., Cham, T.J. & Murphy, K.P. (1999). A Dynamic Bayesian Network Approach to Figure Tracking Using Learned Dynamic Models. *International Conference on Computer Vision*. Corfu, Greece, September, 1999.

Plankers, R.; Fua, P. & D'Apuzzo, N. (1999). Automated Body Modeling from Video Sequences. *International Workshop on Modeling People at ICCV'99*. Corfu, Greece, September, 1999.

Rohr, K. (1997). *Human Movement Analysis Based on Explicit Motion Models*. Kluwer Academic Publishers, Dordrecht Boston, 1997.

Ross, S.M. (1987). *Introduction to Probability and Statistics for Engineers and Scientists*. Wiley Series in Probability and Mathematical Statistics. 1987.

Segawa, H.; Shioya, H., Hiraki, N. & Totsuka, T. (2000). Constraint-Conscious Smoothing Framework for the Recovery of 3D Articulated Motion from Image Sequences. *The fourth International Conference on Automatic Face and Gesture Recognition*. Grenoble, France, March, 2000.

Sidenbladh, H.; De la Torre, F. & Black, M.J. (2000). A Framework for Modeling the Appearance of 3D Articulated Figures. *The fourth International Conference on Automatic Face and Gesture Recognition*. Grenoble, France, March, 2000.

Sminchisescu, C. (2002). Consistency and Coupling in Human Model Likelihoods. *International Conference on Automatic Face and Gesture Recognition.* Washington D.C.

Soslowsky, L.J.; Flatow, E.L., Bigliani, L.U. & Mow, V.C. (1992). Articular Geometry of the Glenohumeral Joint. *Journal on Clinical Orthopaedics and Related Research*. Vol. 285, 1992.

Wachter, S. & Nagel, H.-H. (1999). Tracking Persons in Monocular Image Sequences. *Journal on Computer Vision and Image Understanding*. Vol. 74, nr. 3, 1999.

Wren, C.R.; Clarkson, B.P. & Pentland, A.P. (2000). Understanding Purposeful Human Motion. *The fourth International Conference on Automatic Face and Gesture Recognition*. Grenoble, France, March, 2000.

Wu, Y.; Hua, G. & Yu, T. (2003). Tracking Articulated Body by Dynamic Markov Network. *International Conference on Computer Vision*, Nice, France. 2003.

Zatsiorsky, V.M. (1998). *Kinematics of Human Motion*. Champaign, IL: Human Kinetics, 1998.

Cutting Edge Robotics

Edited by Vedran Kordic, Aleksandar Lazinica and Munir Merdan

ISBN 3-86611-038-3

Hard cover, 784 pages

Publisher Pro Literatur Verlag, Germany

Published online 01, July, 2005

Published in print edition July, 2005

This book is the result of inspirations and contributions from many researchers worldwide. It presents a collection of wide range research results of robotics scientific community. Various aspects of current research in robotics area are explored and discussed. The book begins with researches in robot modelling & design, in which different approaches in kinematical, dynamical and other design issues of mobile robots are discussed. Second chapter deals with various sensor systems, but the major part of the chapter is devoted to robotic vision systems. Chapter III is devoted to robot navigation and presents different navigation architectures. The chapter IV is devoted to research on adaptive and learning systems in mobile robots area. The chapter V speaks about different application areas of multi-robot systems. Other emerging field is discussed in chapter VI - the human- robot interaction. Chapter VII gives a great tutorial on legged robot systems and one research overview on design of a humanoid robot.The different examples of service robots are showed in chapter VIII. Chapter IX is oriented to industrial robots, i.e. robot manipulators. Different mechatronic systems oriented on robotics are explored in the last chapter of the book.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821