

Applying High-Level Knowledge to Obtain an Efficient State-Space Representation of Body Parameters

Thomas B. Moeslund
 Laboratory of Computer Vision and Media Technology
 Aalborg University, Denmark
 E-mail: tbm@cvmt.dk

December 2, 2003

Abstract

Computer vision-based tracking of body parameters is often carried out by model-based approaches as they allow an easy incorporation of high-level knowledge. However, due to a normally large solution space an exhaustive search is not possible and therefore methods based on prediction are applied. To avoid the inherent problems related to prediction we in this paper suggest to apply additional high-level knowledge to represent the state-space more efficiently and hereby allow for an exhaustive search. Concretely, the paper presents a novel representation of the human arm and shows how to prune the state-space of this representation by including high-level knowledge in the form of kinematic constraints. Results show that our pruned state-space contains $1.3 \cdot 10^3$ different configuration on average. In comparison, the standard representation of the arm contains $8.28 \cdot 10^8$ different configuration, hence we have obtained a more efficient representation.

1. Introduction

Computer vision-based tracking of human body parameters is an important problem due to the vast amount of applications [11]. The human body is a self-occluding articulated object which makes the problem difficult. High-level knowledge is therefore included into the algorithms in order to simplify matters. Concretely, model-based approaches are widely used as they allow for a direct incorporation of high-level knowledge in the form of the possible configuration of the object, see [11] for an overview.

Model-based methods have the problem of a large solution space which inhibits an exhaustive search. That is, as the solution space in a model-based approach is spanned by the number of parameters used to model the object, the total number of different solutions are equal to $\prod \lambda_i$, where λ_i is the resolution of the i 'th parameter used to model the object. For example, if we model a piece of a line which can rotate 360° in the image plane with a resolution of 1°

and scale 200 pixels with a resolution of one pixel, then we have a solution space containing $7.20 \cdot 10^4$ different configurations. Even in this simple example we already push the limit of what an exhaustive search can achieve in real-time.

Three standard solutions exists, which avoids the exhaustive search; prediction followed by either an iterative search, a Kalman Filter, or an exhaustive search in the proximity of the prediction. They are all based on the assumption that the object do not change dramatically between two consecutive frames and hence, a nearly correct prediction can be provided. This assumption is in general true, but a good prediction requires the true configuration to be found in the previous frame, and this is not always possible due to e.g., noise and occlusions. Furthermore, in the first frame no prediction can be provided. For articulated objects with self-occlusion the solution space will in general contain multiple modes, i.e., local extrema exist. The above mentioned problems are probably the main reasons why many model-based systems require special environments in order to successfully operate for a longer period of time.

In this paper we will show that an alternatively approach is possible, namely to include additional high-level knowledge into a model-based tracking framework and thereby making it more realistically to apply an exhaustive search. The focus in this paper is how to apply high-level knowledge to more efficiently model the object, i.e., the state-space, and how to prune this state-space. The context is to track the 3D pose of a human arm given a monocular camera setup. As the tracking of the arm will often be a sub-problem of tracking the entire human pose, we assume some other algorithm has already estimated the main body pose, i.e., we know the position of the shoulder.

1.1. Outline of the Paper

In section 2 we first describe the Euler's angles representation, which is the standard way of modelling the arm. We then describe how the Euler's angles representation can be enhanced by applying more high-level knowledge in the

form of kinematic constraints. In section 3 we present the local screw axis model, which is a novel method for modelling the arm. As opposed to the four-dimensional state-space in the Euler's angles representation this modelling scheme only requires two parameters to span the state-space. In sections 4 and 5 we show how to prune the state-space of the local screw axis model by including both dynamic and static constraints. In section 6 the results are presented and discussed in section 7.

2. Euler's Angles Representation

The standard way of representing the state-space of the arm is to apply four Euler's angles, see figure 1.A. Given an unconstrained Euler's angles representation, the size¹ of the state-space is $(360^4 =) 1.68 \cdot 10^{10}$. As many of these configurations are physical impossible the static kinematic constraints of the joint angles can be applied to prune the state-space, see e.g., [2, 9, 12, 13].

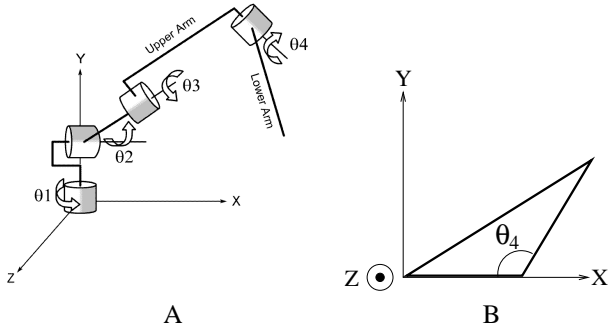


Figure 1: A: The arm represented by four Euler's angles in a coordinate system with origin in the shoulder. The rotation direction of the Euler's angles are indicated by the arrows. B: The reference configurations for the first three Euler's angles, i.e., when the angles are zero. Note that the triangle only lies in the X-Y-plane

The static constraints on the joint angles can be found in many books on anatomy. However, no universal limits exist and the figures differ between the different books. This is due to the fact that the limits differ between individuals and by the actions carried out. As a compromise we have measured the limits for the first author and used these in this work. The ranges are listed in table 1.

Applying the static limits prune the state-space to $(180 \cdot$

¹By 'size' we mean the number of different configurations that can be represented by the state-space. This requires a known resolution of the axes that span the state-space and in this work we use a resolution of $1cm$ for Cartesian variables and 1° for angular variables. Furthermore, in order to present concrete numbers for the pruning effects of the different constraints we assume a framerate of $10Hz$ and that the lengths of the upper arm and lower arm are $A_u = 30cm$ and $A_l = 30cm$, respectively.

Table 1: The legal ranges of the four Euler's angles.

| | θ_1 | θ_2 | θ_3 | θ_4 |
|---------|--------------|--------------|-------------|-------------|
| Minimum | -135° | -135° | 0° | 45° |
| Maximum | 45° | 100° | 145° | 180° |

$235 \cdot 145 \cdot 135 =) 8.28 \cdot 10^8$ different configurations. A significant pruning, but still a large solution space.

Below we show how the state-space can be pruned by applying higher order kinematic constraints, i.e., the velocity and acceleration, but first we need to define the ranges for these constraints. Their ranges depend on the activity carried out. For example, during hammering velocities up to $1031^\circ/s$ and accelerations up to $23778^\circ/s^2$ have been reported [4], and for professional baseball pitchers velocities above $8000^\circ/s$ have been reported [6, 7]. In this work we consider "normal" activities and base our limits on the work by Au and Kirsch [1]. They present measurements of the velocities and accelerations during three "normal" activities: serial single joint movements, reaching movements, and drawing movements. From their results we find the maximum velocity to be $400^\circ/s$ and the maximum acceleration to be $= 4000^\circ/s^2$.

2.1. Including Higher Order Kinematics

The calculations of how higher order kinematics can limit the ranges of the four joint angles are similar and therefore only shown for the general case, θ . The calculations for θ are divided into two parts those results yield $\Delta\theta$, i.e., the maximum displacement of θ . In the first part the maximum change in the positive rotation direction, denoted $\Delta^+\theta$, is found. The second part concerns the maximum change in the negative direction, $\Delta^-\theta$. The calculations are first carried out for $\Delta^+\theta$.

The task at hand is to calculate the maximum displacement of θ in the positive direction, i.e., $\Delta^+\theta$, during one frame given the limits on the angle velocity, V_{max} , and angle acceleration, A_{max} . The displacement in the positive direction is in general calculated as

$$\Delta^+\theta(t) = \frac{1}{2}A_0 \cdot t^2 + V_0 \cdot t \quad (1)$$

where $\Delta^+\theta(t) = \theta(t) - \theta_0$ and θ_0 , A_0 , and V_0 are the angle, acceleration, and velocity at $t = 0$, respectively.

The maximum displacement in the positive direction during one frame is found as

$$\Delta^+\theta = \frac{1}{2} \frac{A_{max}}{f^2} + \frac{V^*}{f} \quad (2)$$

where V^* is the initial velocity. Since the velocity is limited, equation 2 only apply as long as the current velocity,

V , is less than the maximum velocity, i.e., $V < V_{max}$. To evaluate this, the time, t_1 , it takes to reach the maximum accelerate is found as

$$V_{max} = A_{max} \cdot t_1 + V^* \Rightarrow t_1 = \frac{V_{max} - V^*}{A_{max}} \quad (3)$$

After this time the acceleration is zero and the displacement is found as

$$\Delta^+\theta = V_{max}(1/f - t_1) \quad (4)$$

which only apply when

$$t_1 < \frac{1}{f} \Rightarrow V^* > V_{max} - \frac{A_{max}}{f} \quad (5)$$

The displacement can now be expressed as

$$\Delta^+\theta = \begin{cases} \frac{1}{2} \frac{A_{max}}{f^2} + \frac{V^*}{f} & \text{if } V^* \leq V_{max} - \frac{A_{max}}{f} \\ \frac{1}{2} A_{max} t_1^2 + V^* t_1 + V_{max}(1/f - t_1) & \text{if } V^* > V_{max} - \frac{A_{max}}{f} \end{cases} \quad (6)$$

Writing equation 6 as a function of V^* yields

$$\Delta^+\theta(V^*) = \begin{cases} \frac{1}{f} V^* + \frac{1}{2} \frac{A_{max}}{f^2} & \text{if } V^* \leq V_{max} - \frac{A_{max}}{f} \\ \frac{-1}{2A_{max}} \cdot V^{*2} + \frac{V_{max}}{A_{max}} \cdot V^* + \frac{V_{max}}{f} - \frac{V_{max}^2}{2A_{max}} & \text{if } V^* > V_{max} - \frac{A_{max}}{f} \end{cases} \quad (7)$$

Inserting the values from footnote 1, $V_{max} - A_{max}/f = 0$, i.e., the first expression of equation 7 is used when $V^* \leq 0$ and the other one when $V^* > 0$. Altogether the maximum displacement as a function of V^* is shown as the topmost curve in figure 2.

The maximum displacement in the negative direction, $\Delta^-\theta$, can be calculated in a similar manner and the result is illustrated as the lower curve in figure 2. The combined displacement, $\Delta\theta$, as a function of V^* is defined as the area between the two curves in figure 2. This is denoted the relative displacement of θ . The absolute displacement defines the allowed interval for θ and can be found as

$$\Phi = [\Phi_{min}, \Phi_{max}] = [\max\{\theta_{min}, \theta^* + \Delta^-\theta\}, \min\{\theta_{max}, \theta^* + \Delta^+\theta\}] \quad (8)$$

where θ^* is the value of the joint angle in the previous frame, and θ_{min} and θ_{max} are defined in table 1. The minimum and maximum values of θ in the current frame are

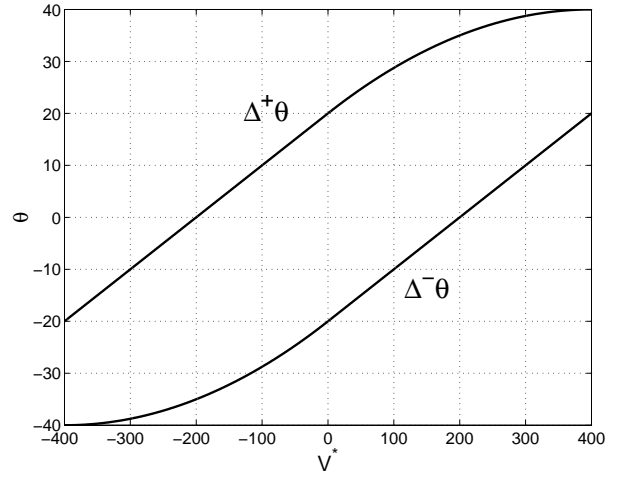


Figure 2: The upper and lower bounds of the displacements.

denoted Φ_{min} and Φ_{max} , respectively. In some situations, e.g., when $\theta^* = \theta_{max} \wedge V^* > 200^\circ/s$ a sudden stop of the movement of the hand is required to obey the kinematic constraints. Obviously, this is not realistic and a minimum interval wherein the hand is bound to be is therefore defined to be equal to the minimum value of $\Delta\theta$, i.e., 20° . Equation 8 is therefore expanded to

$$\Phi = \begin{cases} [\theta_{max} - 20, \theta_{max}] & \text{if } (\Phi_{max} - \Phi_{min}) < 20 \\ & \wedge \Phi_{min} > \theta_{min} \\ [\theta_{min}, \theta_{min} + 20] & \text{if } (\Phi_{max} - \Phi_{min}) < 20 \\ & \wedge \Phi_{max} < \theta_{max} \\ [\Phi_{min}, \Phi_{max}] & \text{otherwise} \end{cases} \quad (9)$$

Expression 9 and figure 2 clearly show the effect of including the higher order kinematic constraints in the pruning. In fact, in worst case the size of the solution space is now $(40^4 =) 2.56 \cdot 10^6$ which is a factor 323 smaller compared to using merely the constraints listed in table 1.

3. Local Screw Axis Model

Even when the velocity and acceleration limits are considered, the resulting solution space is still too large for an exhaustive search. In this section we therefore describe an alternative approach to modelling the arm and show how this novel model can be pruned.

An alternative to the Euler's angles representation is to apply the screw axis representation. It is not directly related to the anatomic joints, but nevertheless has the same ability to represent all the different arm configurations as the Euler's angles representation has. This representation is applied in robotics and computer graphics and recently also in the computer vision community [3] [8].

The representation is based on Chasles' theorem [14] which 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, α , of an initial elbow position around the screw axis [10]. As 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 of α and can be calculated without adding additional parameters. Altogether the representation requires four parameters. Three for the position of the hand, H_x , H_y , and H_z , to define the screw axis and one for the rotation around the screw axis, α .

In order to achieve a smaller solution space we introduce two assumptions. Firstly, we assume that we are working with a calibrated camera. Secondly, we assume we can always locate the hand in the image by utilising skin-color segmentation. Combining the two we can map the position of the centroid of the hand found in the image to a line, l , in space passing through the hand.

$$\vec{H}(t) = \vec{P} + t \cdot \vec{D} \quad (10)$$

where \vec{P} is the focal point of the camera and \vec{D} is a unit direction vector. In other words, given the position of the hand in the image, we only require one free parameter to represent its 3D position, namely the displacement, t . For each value of t the position of the hand \vec{H} is defined. This means that we might as well use any of the entries in \vec{H} to represent the free parameter. In this work we use H_z as the free parameter. So, for each value of H_z t and therefore also H_x and H_y are uniquely determined. By applying this concept to the screw axis representation we can eliminate the parameters H_x and H_y which leave us with just two parameters, namely α and H_z , to model the configuration of the arm. We denote this novel model the *local screw axis model*. For each new image a unique instance of the solution space exists, hence the name "local".

In the local screw axis model α is bounded by one circle-sweep ($0^\circ - 360^\circ$) while H_z is bounded by \pm the total length of the arm. Given the before mentioned values, see footnote 1, we have a solution space containing $4.32 \cdot 10^4$ different solutions. A large reduction in the size compared to that produced by Euler's angles, but still a huge space. The rest of this paper is therefore devoted to applying constraints to prune the local screw axis model. The pruning falls in two categories, four constraints that prunes H_z and two constraints that prunes α .

4. Pruning H_z

H_z is pruned first by three static constraints and then by one dynamic constraint.

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 the hand position is limited by a sphere centred in the shoulder. To calculate the allowed interval for H_z we find where equation 10 intersects the sphere, yielding $P_z + t_2 \cdot D_z \leq H_z \leq P_z + t_1 \cdot D_z$

The second constraint is an angle constraint. When the position of the hand in the image is to the left of the shoulder ($H_x > 0$), see figure 1.A, H_z can be pruned using the limits on θ_1 . The limitations on the angle means, among other things, that the upper arm can only rotate 45° backwards. Together with the limitations of the other angles the minimum H_z positions of the hand is bound to be limited by θ_1 in the following way: $\tan(45^\circ) \geq \frac{H_z}{H_x}$, where the angle is measured from the X-axis. Inserting equation 10 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 equation 10 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 $i = k$ [10]. Equation 10 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 $H_{z,min} = P_z + t \cdot D_z$.

4.1. Dynamic Constraint for H_z

The task at hand is to find the largest and smallest value of H_z , given the higher order kinematics and the values of θ_1 and θ_4 in the previous frame², θ_1^* and θ_4^* .

Using the angles illustrated in figure 3 yields

$$H_z(\theta_1, \theta_4) = -1(\sin(\theta_1) \cdot A_u + \sin(\varphi_2) \cdot A_l) \quad (11)$$

where $\varphi_2 = \theta_1 - 180^\circ + \theta_4$. The change of sign is necessary due to the definition of the rotation directions, see figure 1. The minimum and maximum values of $H_z(\theta_1, \theta_4)$ are found by partial differentiation of H_z and setting both expressions equal to zero. That is

²We ignore θ_2 and θ_3 as we have been unable to find a simple solution if all four angles are considered. We here present a conservative solution by assuming that all movements take place in the X-Z-plane.

$$\frac{\partial}{\partial \theta_1} H_z(\theta_1, \theta_4) = 0 \Rightarrow$$

$$-\cos \theta_1 \cdot A_u + \cos(\theta_1 + \theta_4) \cdot A_l = 0 \quad (12)$$

$$\frac{\partial}{\partial \theta_4} H_z(\theta_1, \theta_4) = \cos(\theta_1 + \theta_4) \cdot A_l = 0 \quad (13)$$

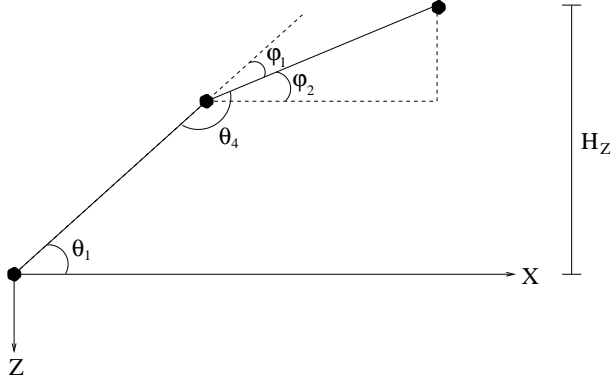


Figure 3: The angles used to calculate the average pruning effect.

It follows that global extrema are found when $\theta_1 = \pm 90^\circ \wedge \theta_4 = 180^\circ$. This intuitive result means that the global extrema are either outside or on the borders of the legal region defined by the limits of θ_1 and θ_4 , see expression 9. Only saddle points may be found inside the region meaning that the minimum and maximum values of H_z always will be located on the borders. The four borders are therefore investigated one at the time and the minimum and maximum values are found.

For the border where θ_4 is fixed to its maximum value, i.e., $\theta_4 = \Phi_{4,max}$, the following equation is solved

$$\frac{d}{d\theta_1} H_z(\theta_1, \Phi_{4,max}) = 0 \Rightarrow$$

$$-\cos(\theta_1) \cdot A_u + \cos(\theta_1 + \Phi_{4,max}) \cdot A_l = 0 \quad (14)$$

The equation can be rewritten to the transcendental equation which has two solutions $\theta_{1,1}$ and $\theta_{1,2}$ [10]. If the solutions are within the legal interval of θ_1 for the current frame ($[\Phi_{1,min}, \Phi_{1,max}]$) the minimum value, σ_1 , and maximum value, ω_1 , for the border are found as

$$\sigma_1 = \min\{H_z(\theta_{1,1}, \Phi_{4,max}), H_z(\theta_{1,2}, \Phi_{4,max}),$$

$$H_z(\Phi_{1,min}, \Phi_{4,max}), H_z(\Phi_{1,max}, \Phi_{4,max})\} \quad (15)$$

$$\omega_1 = \max\{H_z(\theta_{1,1}, \Phi_{4,max}), H_z(\theta_{1,2}, \Phi_{4,max}),$$

$$H_z(\Phi_{1,min}, \Phi_{4,max}), H_z(\Phi_{1,max}, \Phi_{4,max})\} \quad (16)$$

Note that the values of the corners of the legal interval is included to handle the situation where less than two local extrema can be found.

The minimum and maximum for the three other borders are found in a similar manner. The overall minimum and maximum values can now be found as

$$\min\{H_z(\theta_1, \theta_4)\} = \min\{\sigma_1, \sigma_2, \sigma_3, \sigma_4\} \quad (17)$$

$$\max\{H_z(\theta_1, \theta_4)\} = \max\{\omega_1, \omega_2, \omega_3, \omega_4\} \quad (18)$$

Since this constraint is calculated without taking θ_2 and θ_3 into consideration, the intervals in 17 and 18 are relative with respect to the actually value of H_z in the previous frame, H_z^* . The pruning of H_z is therefore given as

$$H_z^* - \min\{H_z(\theta_1, \theta_4)\} \leq H_z \leq H_z^* + \max\{H_z(\theta_1, \theta_4)\} \quad (19)$$

5. Pruning α

In this section we will describe two constraints that prune α . First a collision constraint and then a constraint based on the limits on the joint angles.

The fact that two body parts can not occupy the same space at the same time provides a collision constraint. This constraint is applied for each non-pruned H_z -value and it is therefore important to have an analytic solution which evaluate 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 rectangle in the X-Z-plane [10]. For each value of H_z the elbow describes a circle in space and, in general, an ellipse in the X-Z-plane. 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).

5.1. Pruning α using the Joint Angle Limits

The task at hand is for a certain value of H_z , i.e., \vec{H} , to find a relationship between the four joint angles and α , so that the limitations from section 2 may be applied to limit the range of α .

First of all it is noted that θ_4 can be calculated for each (H_z, α) set and needs to be within $[\Phi_{4,min}, \Phi_{4,max}] = [45^\circ, 180^\circ]$. If not, the current value of H_z can be ignored altogether.

The approach used to find a relationship between the remaining three angles and α is based on the idea of explaining how to rotate the reference triangle in figure 1.B

into the current triangle. The rotation can be described by $\mathbf{R}(\theta_1, \theta_2, \theta_3)$ and by $\mathbf{R}(\alpha)$. As the two rotations will be equal a relationship between the joint angles and α is given by $\mathbf{R}(\theta_1, \theta_2, \theta_3) = \mathbf{R}(\alpha)$.

The former rotation is an Y-Z-X Euler's angles rotation, i.e., $\mathbf{R}(\theta_1, \theta_2, \theta_3) = \mathbf{R}_y(\theta_1) \cdot \mathbf{R}_z(\theta_2) \cdot \mathbf{R}_x(\theta_3)$ where $\mathbf{R}_y(\theta_1)$ is the rotation of θ_1 around the Y-axis etc.

5.1.1 Finding $\mathbf{R}(\alpha)$

Five rotations are used to align the reference triangle, see figure 1.B, with the current triangle, i.e., pose of the arm. The first rotation is to align the shoulder-hand-line of the reference triangle, \vec{H}_r , with the X-axis. The rotation is done using $\mathbf{R}_z(\phi_1)$ where

$$\phi_1 = -\sin^{-1} \left(\frac{A_l \sin(\theta_4)}{\|\vec{H}_r\|} \right) \quad (20)$$

and $\|\vec{H}_r\| = \|\vec{H}\|$

Next, the triangle is rotated 180° about the X-axis. The third and fourth rotations align the shoulder-hand-line of the reference triangle with the current shoulder-hand-line, \vec{H} . This is done by first rotating around the Z-axis and then around the Y-axis by $\mathbf{R}_z(\phi_2)$ and $\mathbf{R}_y(\phi_3)$, respectively. The rotation angles are found as $\phi_2 = \sin^{-1} \left(\frac{H_y}{\|\vec{H}\|} \right)$ and $\phi_3 = \tan^{-1} \left(\frac{H_z}{H_x} \right)$.

The shoulder-hand-line of the reference triangle is now aligned with the shoulder-hand-line of the current configuration, i.e., \vec{H}_r is aligned with \vec{H} . Finally the position of the elbow in the reference triangle needs to be aligned with the current elbow position. This is done by rotating the reference triangle around \vec{H} with α degrees. For this purpose the Rodrigues' formula [5] is applied. It rotates a vector \vec{P} α degrees about a unit vector \vec{K} to a new vector \vec{P}'

$$\vec{P}' = \vec{P} \cos(\alpha) + \sin(\alpha)(\vec{K} \times \vec{P}) + (1 - \cos(\alpha))(\vec{K} \cdot \vec{P})\vec{K} \quad (21)$$

The unit vector is given as $\vec{K} = \frac{\vec{H}}{\|\vec{H}\|}$ and \vec{P}' is the current position of the elbow \vec{E} . \vec{P} is the top point (highest Y-value) of the current α -circle and is denoted \vec{E}_T . Due to the rotation, $\mathbf{R}_x(180^\circ)$, the elbow of the reference triangle is always in the topmost position (with respect to y). Equation 21 can be written in matrix form as $\vec{E} = \mathbf{R}_e(\alpha) \cdot \vec{E}_T$ [10].

The overall rotation needed to align the reference triangle with the current triangle in terms of α is given as

$$\mathbf{R}(\alpha) = \mathbf{R}_e(\alpha) \mathbf{R}_y(\phi_3) \mathbf{R}_z(\phi_2) \mathbf{R}_x(180^\circ) \mathbf{R}_z(\phi_1) \quad (22)$$

5.1.2 Relating $(\theta_1, \theta_2, \theta_3)$ and α

After $\mathbf{R}(\theta_1, \theta_2, \theta_3)$ and $\mathbf{R}(\alpha)$ are found the relationship between the three joint angles and α can be found. As stated above $\mathbf{R}(\theta_1, \theta_2, \theta_3) = \mathbf{R}(\alpha)$. For each of the nine entries in the matrices a trigonometric equation relating the joint angles and α exists. We apply the entries shown below to find a solution.

$$\sin(\theta_2) = a_1 \cdot \sin(\alpha) + b_1 \cdot \cos(\alpha) + c_1 \quad (23)$$

$$\cos(\theta_1) \cos(\theta_2) = a_2 \cdot \sin(\alpha) + b_2 \cdot \cos(\alpha) + c_2 \quad (24)$$

$$\cos(\theta_2) \cos(\theta_3) = a_3 \cdot \sin(\alpha) + b_3 \cdot \cos(\alpha) + c_3 \quad (25)$$

where a_i , b_i , and c_i are found from \vec{H} , and therefore constant for each value of H_z .

Relating θ_2 and α

Finding the legal set of α in accordance with the limits on the three joint angles is done by first finding the legal sets for each joint angle and then combining them.

First, it is calculated how θ_2 limits α using equation 23. The legal interval of θ_2 is $[\Phi_{2,min}, \Phi_{2,max}]$ as defined in 2.1. All legal values of α have to satisfy equation 23 so that $\theta_2 \in [\Phi_{2,min}, \Phi_{2,max}]$. In figure 4 this is illustrated. The sine-curve is the right hand-side of equation 23 and the straight lines are minimum- and maximum values of the left hand-side of equation 23. For an α value to be accepted $t(\alpha)$ needs to be within the shaded area. The shaded parts of the α -axis indicate the legal intervals of α .

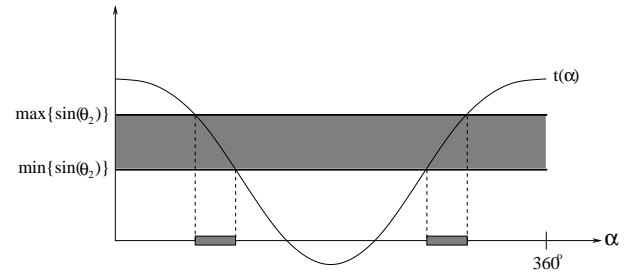


Figure 4: Illustrating the relationship between θ_2 and α .

Due to the monotonic characteristics of $\sin(\theta_2)$ the estimation of the legal intervals are split up into three parts, one for $\theta_2 \in]-270^\circ, -90^\circ]$ one for $\theta_2 \in]-90^\circ, 90^\circ]$ and one for $\theta_2 \in]90^\circ, 270^\circ]$. These intervals are denoted p_1 , p_2 , and p_3 , respectively. Combining p_1 and $[\Phi_{2,min}, \Phi_{2,max}]$, p_2 and $[\Phi_{2,min}, \Phi_{2,max}]$, and p_3 and $[\Phi_{2,min}, \Phi_{2,max}]$, the legal intervals used to define the shaded area in figure 4 are found as $l_1 = [\Phi_{2,min}, \Phi_{2,max}] \cap p_1$, $l_2 = [\Phi_{2,min}, \Phi_{2,max}] \cap p_2$, and $l_3 = [\Phi_{2,min}, \Phi_{2,max}] \cap p_3$, respectively. Within these sets $\sin(\theta_2)$ is defined as

$$l_1 : \sin(\theta_2) \in [\sin(\max\{l_1\}), \sin(\min\{l_1\})] \quad (26)$$

$$l_2 : \sin(\theta_2) \in [\sin(\min\{l_2\}), \sin(\max\{l_2\})] \quad (27)$$

$$l_3 : \sin(\theta_2) \in [\sin(\max\{l_3\}), \sin(\min\{l_3\})] \quad (28)$$

In the following the calculations for l_1 are shown. In theory three different solution types exist when solving equation 23. The first is when the shaded area do not intersect with $t(\alpha)$ meaning that the arm cannot be in the configuration defined by the current value of H_z . In this case no further calculations are necessary, hence 100% pruning of α . In the second case α is not constrained by θ_2 , i.e., θ_2 do not prune α . This occur when all of $t(\alpha)$ is inside the shaded area. In the third case the lines and the curves intersect. To decide whether they intersect the minimum and maximum values of $t(\alpha)$ is found and compared with the boundaries of the shaded area.

If at least one boundary intersects the sine-curve, the intersection points are found by solving equation 23 for $\sin(\theta_2) = \sin(\min\{l_1\})$ and $\sin(\theta_2) = \sin(\max\{l_1\})$, respectively. Since the left-hand side of both equations are constant the equation can be re-written to transcendental equations each with two solutions [10]. For each equation zero, one, or two intersection points are found, hence between one and four α values altogether. These are together with the two extrema of α (0 and 360°) placed in a list and sorted according to their magnitude. Depending on the signs and magnitude of the coefficients in equation 23 and the values of α , the interval between two consecutive values in the list is either legal or illegal. Instead of setting up rules it is easier to use the monotonic characteristics of the curve and evaluating one point in each interval. If and only if this point is legal, the interval is legal. In other words

$$\alpha \in [\alpha_i, \alpha_{i+1}] \quad \text{if } t((\alpha_i + \alpha_{i+1})/2) \in [\sin(\min\{l_1\}), \sin(\max\{l_1\})] \quad (29)$$

where α_i is the i 'th value in the sorted list.

Altogether the union of the legal intervals defines the legal set of α with respect to l_1 . Similar calculations are carried out for l_2 and l_3 and the union of the three results defines the legal set of α -values with respect to θ_2 . This set is denoted Π_{θ_2}

Relating θ_1 and α

To calculate how θ_1 limits α equation 24 is applied. Since two joint angles are present in this equation, the legal intervals l_i are defined over two variables rather than one, hence a legal region. This can result in more intervals due to more complex monotonic characteristics, but otherwise the calculations are similar to those explained above.

In order to find the legal regions the monotonic characteristics are investigated. This is carried out by calculating the partial derivatives of the left hand side of equation 24 and setting them equal to zero. This yields local extrema for each variable along the lines $\theta_1 = -90^\circ$, $\theta_1 = 0^\circ$, $\theta_2 = -90^\circ$, $\theta_2 = 0^\circ$, and $\theta_2 = 90^\circ$. The extrema for the combined variables, i.e., $\cos(\theta_1) \cos(\theta_2)$, are located where the lines intersect. Together with the legal intervals of the two joint angle, $[\Phi_{1,min}, \Phi_{1,max}]$ and $[\Phi_{2,min}, \Phi_{2,max}]$, the lines form a grid wherein each rectangle constitutes a legal region, l_i .

Within each legal region the value of the left hand side of equation 24 is monotone with minimum and maximum defined as

$$l_i : \cos(\theta_1) \cdot \cos(\theta_2) \in [\min\{c_{i,1}, c_{i,2}, c_{i,3}, c_{i,4}\}, \max\{c_{i,1}, c_{i,2}, c_{i,3}, c_{i,4}\}] \quad (30)$$

where $c_{i,1} \dots c_{i,4}$ are the values of the four corners defining the i 'th legal region, e.g., $c_{i,1} = \cos(0^\circ) \cdot \cos(0^\circ) = 1$.

Having found the legal regions, l_i , and their maximum and minimum values, the calculations are similar to those described above. The result is a set of legal α -values with respect to θ_1 and θ_2 . This set is denoted Π_{θ_1} .

5.1.3 General Pruning Effect of θ_1 , θ_2 , and θ_3

To calculate how θ_3 limits α equation 25 is applied. The calculation of Π_{θ_3} is exactly the same as when calculating Π_{θ_1} . Altogether the three joint angles constrain α to be in the set Π defined as $\Pi = \Pi_{\theta_1} \cap \Pi_{\theta_2} \cap \Pi_{\theta_3}$.

6. Results

The results of this work is the pruning effects of the different constraints.

How much a particular constraint prunes the solution space depends on the current position of the hand and the previously estimated configuration of the arm³. It is therefore not possible to state a general pruning effect but in the upper part of table 2 the intervals of the pruning effects are shown together with the average effects [10].

Since the constraints are dependent it is not possible to calculate a general accumulated effect. Instead the total minimum, maximum, and average effects can be estimated, see the lower part of the table for the numbers and [10] for further information.

If we ignore all constraints the local screw axis model contains $(120 \cdot 360 =) 4.32 \cdot 10^4$ different configurations. Applying the total pruning effects listed in table 2 we have

³The 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 infinitely far away.

Table 2: The different constraints and their pruning effects.

| Par. | Type of constraint | Pruning Effects | | |
|---------------|--------------------------|-----------------|------|------|
| | | Min. | Max. | Avg. |
| H_z | Distance constraint | 0% | 100% | 48% |
| H_z | Angle constraint | 0% | 50% | 25% |
| H_z | Occlusion constraint | 0% | 57% | 20% |
| H_z | Temporal constraint | 51% | 92% | 77% |
| α | Collision constraint | 0% | 100% | 35% |
| α | Joint angle constraint | 75% | 100% | 85% |
| H_z | Total effect on H_z | 51% | 100% | 80% |
| α | Total effect on α | 75% | 100% | 86% |
| α, H_z | Total effect | 88% | 100% | 97% |
| α, H_z | Total effect (static) | 75% | 100% | 92% |

a solution space containing between 0 and $5.18 \cdot 10^3$ distinct configurations and $1.3 \cdot 10^3$ on average. And if the resolution is set to e.g., 2cm for H_z and 5° for α , then we have $5.18 \cdot 10^2$ configurations in worst case and $1.30 \cdot 10^2$ on average.

6.1. Temporal Aspects

All calculations involving temporal information are based on a framerate of 10Hz. If the framerate is increased the pruning effects of the fourth and fifth constraints will improve. If the framerate is decreased the pruning effects will decrease. In situations where the framerate is unknown or changing the results of the fourth and fifth constraint become unreliable. We therefore estimate the total effects in the case we ignore all temporal information, hence we apply $\theta \in [\theta_{min}, \theta_{max}]$ rather than $\theta \in [\Phi_{min}, \Phi_{max}]$ in the fourth and fifth constraints. The results of the combined effects are listed in table 2 (static) [10].

Without the temporal constraints the absolute size of the solutions space will be between 0 and $1.08 \cdot 10^4$ distinct configurations and $3.46 \cdot 10^3$ on average. And if the resolution is set to e.g., 2cm for H_z and 5° for α , then we end up with $1.08 \cdot 10^2$ configurations in worst case and $3.46 \cdot 10^2$ on average.

7. Discussion

In this paper we have shown how high-level knowledge can be applied to reduce the size of the state-space of the arm, both in terms of the dimensionality and in terms of the number of different configurations. Our approach is to use kinematic constraints and our reduction is therefore *not* with respect to some application or training data but very general. In other words, we only prune impossible configurations as opposed to improbable configurations.

Our main contribution is the local screw axis model and the pruning of this representation. The model allows a very efficient representation of the arm which in term may allow for an exhaustive search in the solution space. A second

contribution is section 2.1 which shows how to incorporate velocity and acceleration constraints into the Euler's angles representation, i.e., a contribution independent of the local screw axis model.

One issue that needs to be stressed with respect to the local screw axis model is the assumption that the hand can always be segmented in the image. In situations where this fails one should either apply predicted values (only for a limited amount of frames) or switch to the standard Euler's angles representation or screw axis representation, and continue the tracking in one of these state-spaces until the hand can be detected once again.

References

- [1] A.T.C. Au and R.F. Kirsch. EMG-Based Prediction of Shoulder and Elbow Kinematics in Able-Bodied and Spinal Cord Injured Individuals. *IEEE Transactions on Rehabilitation Engineering*, 8(4), 2000.
- [2] C. Barron and I.A. Kakadiaris. Estimating Anthropometry and Pose from a Single Image. In *Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, June 13-15 2000.
- [3] C. Bregler and J. Malik. Tracking People with Twists and Exponential Maps. In *International Conference on Computer Vision and Pattern Recognition*, Santa Barbara, California, June 1998.
- [4] M.A. Buckley, A. Yardley, G.R. Johnson, and D.A. Carus. Dynamics of the upper limb performance of the tasks of everyday living - a review of the current knowledge base. *Engineering in Medicine - Part H*, 210(4), 1996.
- [5] J.J. Craig. *Introduction to Robotics. Mechanics and Control*. Addison Wesley, second edition, 1989.
- [6] R.F. Escamilla, G.S. Fleisig, N. Zheng, S.W. Barrentine, and J.R. Andrews. Kinematic comparison of 1996 Olympic baseball pitchers. *Journal of Sports Sciences*, 19, 2001.
- [7] G.S. Fleisig, S.W. Barrentine, N. Zheng, R.F. Escamilla, and J.R. Andrews. Kinematic and kinetic comparison of baseball pitching among various levels of development. *Journal of Biomechanics*, 32, 1999.
- [8] A. Hilton. Towards Model-Based Capture of a Persons Shape, Appearance and Motion. In *International Workshop on Modeling People at ICCV'99*, Corfu, Greece, September 1999.
- [9] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman. Human Body Model Acquisition and Motion Capture Using Voxel Data. In F.J. Perales and E.R. Hancock, editors, *AMDO 2002*, LNCS 2492. Springer-Verlag, 2002.
- [10] T.B. Moeslund. *Computer Vision-Based Motion Capture of Body Language*. PhD thesis, Laboratory of Computer Vision and Media Technology, Aalborg University, Denmark, 2003.
- [11] T.B. Moeslund and E. Granum. A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81(3), 2001.

- [12] R. Plänkers and P. Fua. Articulated Soft Objects for Video-based Body Modeling. In *International Conference on Computer Vision*, Vancouver, Canada, July 9-12 2001.
- [13] T.J. Roberts, S.J. McKenna, and I.W. Ricketts. Adaptive Learning of Statistical Appearance Models for 3D Human Tracking. In *British Machine Vision Conference*, Cardiff, UK, 2002.
- [14] V.M. Zatsiorsky. *Kinematics of Human Motion*. Champaign, IL: Human Kinetics, 1998.