Machine Vision and Applications (1998) 10: 223–231

Machine Vision and Applications

Automatic reconstruction of 3D human arm motion from a monocular image sequence

Valentina Filova[1], **Franc Solina**[2], **Jadran Lenarčič**[3]

[1] IBL-Sistemi, Železna c. 18, SI-1000 Ljubljana, Slovenia
[2] University of Ljubljana, Faculty of Computer and Information Science, Tržaška c. 25, SI-1000 Ljubljana, Slovenia
[3] Jožef Stefan Institute, Jamova c. 39, SI-1000 Ljubljana, Slovenia

Received: 26 January 1996 / Accepted: 17 July 1997

Abstract. A model-based approach to reconstruction of 3D human arm motion from a monocular image sequence taken under orthographic projection is presented. The reconstruction is divided into two stages. First, a 2D shape model is used to track the arm silhouettes and second-order curves are used to model the arm based on an iteratively reweighted least square method. As a result, 2D stick figures are extracted. In the second stage, the stick figures are backprojected into the scene. 3D postures are reconstructed using the constraints of a 3D kinematic model of the human arm. The motion of the arm is then derived as a transition between the arm postures. Applications of these results are foreseen in the analysis of human motion patterns.

Key words: Model-based tracking – Motion reconstruction – Robust estimation – Human arm kinematics – Constraint propagation

1 Introduction

The study of human motion has received considerable attention over the past decades in robotics, surgery, rehabilitation, behavior studies and sports. A common approach for quantitative estimation of motion is to attach markers on the skin of subjects at selected anatomic points and to record the subject's motion with cameras. The method is known as moving light display (MLD) [3, 9]. A typical application is the evaluation of an athletes' or robot mechanism's motion. The advantage of this approach is that it requires only basic image processing, however, a severe disadvantage of the method is that it is invasive. The subject must wear markers or other equipment which may be impractical, uncomfortable or constrain the usage of the system to a limited work space.

Recently, new applications such as videophones, teleconferencing, multimedia, human-computer interface design, virtual reality, etc. require the capability of following the movement of the human head, eyes, lips, arms and hands from images. The emphasis is not on the recovery of exact motion but on recognition of motion as a meaningful gesture or activity.

The goal of this investigation was to develop a method for monocular tracking of the human arm in 3D without the use of markers. In particular, we were interested if the kinematic model of the human arm that was at our disposal [14] provided enough constraints to do that. We consider the human arm to be a six degree-of-freedom (d.o.f.) articulated object. We assume that the projection is orthographic and that the arm is moving in front of a dark background. Our approach differs from previous approaches of tracking articulated objects in the sense that tracking is achieved without reconstructing the 3D shape and structure information. First, 2D arm segment axes are extracted from arm silhouettes. Next, 3D postures of the arm are reconstructed by backprojecting the 2D stick figures using the constraints of the kinematic model in each frame separately. The motion of the arm is derived as a transition between individual arm postures.

In the next section, we give an overview of the approach. The third and fourth section is on the 2D modeling of the arm segments. In the fifth section, the 3D reconstruction is presented. The sixth section gives some examples of tracking and reconstruction, followed by discussion and conclusions.

1.1 Related work

One of the earliest systems capable of analyzing 3D human motion based on a human model was developed by O'Rourke and Badler [18]. The model consists of 25 segments defined by about 600 overlapping spheres. Given a sequence of monocular images, the subject is tracked using constraint propagation, high-level prediction and low-level verification and analysis. All computations are conducted in 3D space. The authors reported results of tracking the human arms also in the case of occlusion. Yamamoto and Koshikawa [21] developed a technique to reconstruct the human arm motion from a real image sequence combining 3D geometrical and kinematic models of the arm and the image displacement field. The motion equations are derived by establishing a connection between the changes of the kinematic

Correspondence to: V. Filova, e-mail: IBL@integra.mibo.si

configuration and image spatial and temporal gradient of the wire model points. Gavrila and Davis [7] developed a multi-view approach for tracking of the human upper body. The pose recovery problem is formulated as a search problem and entails finding the pose parameters of a human model for which its synthesized appearance is most similar to the actual appearance of the real human. Goncalves et al. [8] developed a system for monocular tracking of the human arm using the extended Kalman filter. The arm in 3D is modeled with truncated right-circular cones. Ohya and Kishino [17] proposed a passive method for estimating the human posture from multiple cameras. A genetic algorithm is used to estimate the posture parameters. Rehg and Kanade [20] developed a model-based hand-tracking system, called Digit Eyes, that can recover the state of a 27 d.o.f. hand model from gray-scale images. 3D tracking is achieved by using a high image acquisition rate. This decreases the differences between the kinematic configurations in the subsequent images and thus simplifies the tracking of image features. Kuch and Huang [13] developed a hand-modeling and -tracking system for virtual teleconferencing and telecollaboration. A 3D model of the hand is built automatically from three different views of the hand. Tracking is achieved by minimizing the error between the 2D appearance of the model and the image. Kakadiaris, et al. [11] proposed an integrated approach to segmentation, shape and motion estimation of complex articulated objects. Unlike most existing techniques which assume a priori knowledge of the object's parts, they initially assume that the data belongs to a single-part object and fit a single deformable model. As the model deforms to fit the data from subsequent frames they decide when to split the model into two new models. In order to cope with occlusion, they use a Kalman filter to predict the location of the data in the next frame. Some of the most impressive results of tracking human body and action recognition are currently achieved by Pentland et al. [1, 15, 19]. They use 2D blob (3D blobs) features to build connected blob representation of the human. From 2D blobs the 3D geometry for the user head and hands are calculated using an uncalibrated stereo system. The system runs in real time.

2 Overview of the approach

Our approach is based on the following:

1. a six d.o.f. kinematic model that defines all possible spatial configurations of the arm,
2. a 2D shape model that describes the arm's visual appearance,
3. an assumption, that 2D stick figures of the arm can be extracted from each image frame.

2.1 Kinematic model of the arm

The simplified (not anatomically precise) kinematic model of the human arm without the palm (Fig. 1), developed by Lenarčič and Umek [14] with the aim to represent the reachability of the arm, has six revolute degrees of freedom: two in the sternoclavicular joint, three in the glenohumeral joint and one in the elbow joint.

The reference coordinate frame is placed in the sternoclavicular joint. Expressed in terms of the matrix algebra, the 3D vectors of the shoulder $\vec{r_{sh}}$, elbow $\vec{r_{el}}$ and wrist joints $\vec{r_{wr}}$ are as follows:

$$\vec{r_{sh}} = \mathbf{R}_z(\delta_1) \cdot \mathbf{R}_y(\delta_2) \cdot \vec{d_{cl}}, \tag{1}$$

$$\vec{r_{el}} = \vec{r_{sh}} + \mathbf{R}_z(\delta_1) \cdot \mathbf{R}_y(\delta_2) \cdot \mathbf{R}_y(\delta_3) \cdot \mathbf{R}_x(\delta_4) \cdot \mathbf{R}_z(\delta_5).\vec{d_{up}}, \tag{2}$$

$$\vec{r_{wr}} = \vec{r_{el}} + \mathbf{R}_z(\delta_1) \cdot \mathbf{R}_y(\delta_2) \cdot \mathbf{R}_y(\delta_3) \cdot \mathbf{R}_x(\delta_4) \cdot \mathbf{R}_z(\delta_5) \cdot \mathbf{R}_x(\delta_6) \cdot \vec{d_{for}}. \tag{3}$$

Vectors $\vec{d_{cl}}$, $\vec{d_{up}}$ and $\vec{d_{for}}$ contain the lengths of the clavicular segment D_{cl}, the upper arm D_{up} and the forearm D_{for}, while matrices $\mathbf{R}_x(\cdot)$, $\mathbf{R}_y(\cdot)$ and $\mathbf{R}_z(\cdot)$ are standard rotation matrices about the x-, y- and z-axes and define the relative revolute displacements between the segments. The kinematic model also defines the dependencies between the joint angles δ_i (Fig. 1).

2.2 2D shape model of the arm

Based on the above kinematic model, we built a 2D shape model of the human arm. It is a composition of fitted second-order curves to the arm silhouette (Fig. 2a). The intersections of the second-order curves define the lines l_2 and l_3 that partition the image data into parts according to the kinematic model. The head is used as a reference to set the root point R and the shoulder partitioning line l_1. To facilitate the process of tracking, the palm is also included in the model. Where the curve axes intersect the partitioning lines, the local coordinate systems for the upper arm S, the forearm E and the palm W are centered respectively. Line segments that connect the root point R, shoulder point S, elbow point E and wrist point W produce a stick figure of the arm (Fig. 2b).

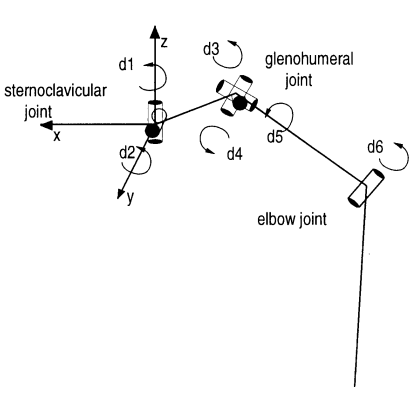
Both models (kinematic and shape) define the state model of the arm which encodes all possible arm configurations and their corresponding image feature patterns. Models give rise to the following **constraints** [10, 12]:

- *Strong constraints*: fixed length of the body parts, a three-point co-linearity of joints, angle relationships, occlusion-collision relationships (defined in the kinematic model).
- *Soft constraints*: two-way correspondence between the image features (region, points) and the 2D shape model.

Assuming *smooth motion* the 3D reconstruction of the arm movements is possible.

2.3 Human arm motion reconstruction method

As the arm moves in 3D its 2D silhouette expresses a variety of shapes. To reconstruct the motion of the arm from a sequence of monocular images, we need to recover the time-varying parameters of the shape model and the kinematic model. We assume that the projection is orthographic and that the initial 3D position of the arm is known. We divided the process of motion tracking and reconstruction into two steps:



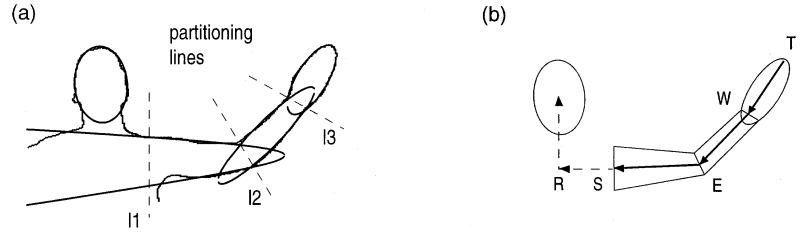
Joint angles:

d1 - clavicular abduction/adduction
d2 - clavicular flection/extension
d3 - humeral flection/extension
d4 - humeral abduction/adduction
d5 - humeral rotation
d6 - elbow flection/extension

Limits of joint angle values [degrees]:

d1 = [-17 , 17]
d2 = [-6 , 20]
d3 = [-9 , 160]
d4 = [-43 + d3/3, 153 - d3/6]
d5 = [-90 + 7 d3/9 - d4/9 + 2 d3 d4/810 ,
60 + 4 d3/9 - 5 d4/9 + 5 d3 d4/810]
d6 = [-90 , 60]

1



2

Fig. 1. 3D kinematic model of the left arm

Fig. 2a. 2D shape model of the arm and **b** its corresponding stick figure

Step 1. In the first step, we track the arm silhouettes in 2D. This is achieved by automatically locating the head and the top of the arm and fitting the 2D shape model to the arm segment projections (Sect. 4). Each segment is modeled separately (Sect. 3). We assume that there is a critical amount of edge points that describe each segment of the arm and that there is no self-occlusion. The result is an extracted 2D stick figure of the arm.

Step 2. In the second step, the stick figures are backprojected into the 3D scene (Fig. 3). The inverse projection problem is solved in each frame separately. The 3D structure recovery equations are derived by propagating the strong constraints of the kinematic model from the root towards the wrist point (Sect. 5). As the projection is assumed to be orthographic, there are two mirror solutions in 3D for each body segment or all together eight possible solutions for the whole arm structure (considering the arm as a three-segment kinematic chain). However, due to the joint angle constraints defined in the kinematic model (Fig. 1), not all solutions are possible. By applying the structure recovery equations in specified order and assuming *smooth motion* (values of the joint angles change smoothly), we get a unique solution for the structure of the arm in 3D. Once the 3D postures are recovered, the 3D motion is given as a transition between two postures in space. The reconstructed 3D motion can be described either with the trajectories of the joint points of the kinematic chain S, E and W or with the time-varying sequences of the joint angles δ_i.

3 Modeling the arm segment projections using robust M-estimation

The appearance of an articulated object such as the human arm on the image plane is fairly complex and difficult to describe with a single model. Standard least square analysis that assumes that the data belongs to a single model and that the errors in the data are normally and identically distributed is of no use. Our aim is to segment the arm silhouettes into distinct parts of the arm and to use these models to determine the 2D joint positions. After much experimentation a second-order curve was chosen to model the data of a single part. An advantage of this model is that the small number of parameters that describe the curve can be directly applied to determine the position and orientation of the arm segments. Taking into consideration that arm segments can appear in different sizes, orientations and shapes, we need a procedure that is robust to local deformations and outliers (data parts from other parts, or just noise), as well as stable to global transformations (rotation, translation and scaling). Robust M-estimation technique based on the *Hampel redescending* function turned out to be good in both aspects [4]. This section proceeds with the mathematical background of the robust parameter estimation technique.

3.1 Mathematical background

The general equation of a second-order curve $f(\vec{p}, x, y)$ is a linear combination of five independent basis functions $\{\phi_1, \phi_2, \phi_3, \phi_4, \phi_5\} = \{x^2, xy, y^2, x, y\}$:

$$f(\vec{p}, x, y) = p_1\phi_1 + p_2\phi_2 + p_3\phi_3 + p_4\phi_4 + p_5\phi_5 + 1. \quad (4)$$

The modeling process can be defined as the search for such a parameter vector $\vec{p} = \{p_1, p_2, p_3, p_4, p_5\}$ of the model $\hat{f}(\vec{p}, x, y)$, which fits best the structure of the image data $U = \{(x_1, y_1), ., (x_n, y_N)\}$.

A robust M-estimate for the parameter vector $\vec{p}$ minimizes the error function $\epsilon(\vec{p})$ that sums the deviations of

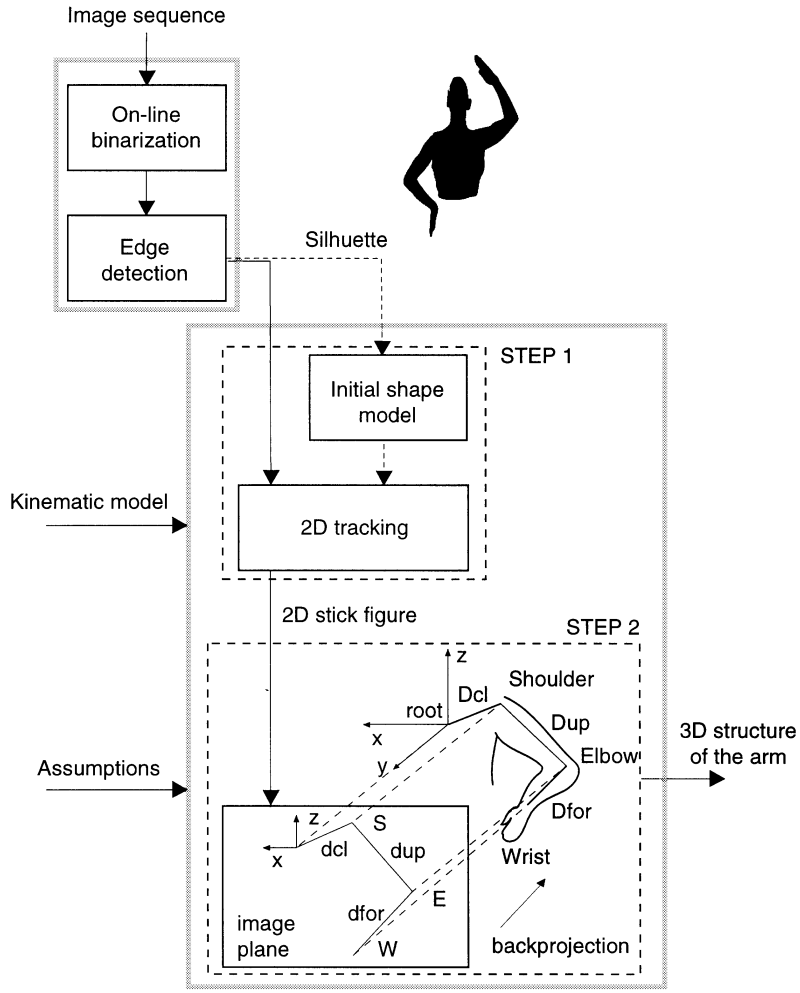


Fig. 3. Overview of the reconstruction approach

the observations $e(x_i, y_i)$ from the fitted curve ($e(x_i, y_i)$ = $f(\vec{p}, x, y)$ = 0 for the points on the curve):

$$\epsilon(\vec{p}) = \sum_{i=1}^{N} \rho(\frac{e(x_i, y_i)}{s}). \quad (5)$$

$$s(\vec{p}) = 1.4826 \;\; median(|e(x_i, y_i) - median(e(x_i, y_i))|). \quad (6)$$

The parameter s is a known or previously computed scale parameter and ρ is a robust loss function. This is more general than the sum of squared deviations ($\rho(x) = x^2$), or the sum of absolute deviations ($\rho(x) = |x|$).

If we let $\psi(\vec{p}, x, y) = \frac{\partial(\rho(\vec{p}, x, y))}{\partial(\vec{p})}$, then a necessary condition for a minimum of the function $\epsilon(\vec{p})$ is that $\vec{p}$ satisfies

$$\sum_{i=1}^{N} \psi(\frac{e(x_i, y_i)}{s}) \quad \phi_m(x_i, y_i) = 0, \qquad m = 1, 2, ., 5. \quad (7)$$

Introducing a set of weighting parameters:

$$\omega(x_i, y_i) = \begin{cases} \frac{\psi(\frac{e(x_i, y_i)}{s})}{\frac{e(x_i, y_i)}{s}} & \text{if } e(x_i, y_i) \neq 0 \\ 1 & \text{if } e(x_i, y_i) = 0 \end{cases} \quad (8)$$

the nonlinear matrix Eq. 7 can be rewritten as follows:

$$\sum_{i=1}^{N} \phi_m(x_i, y_i)\ \omega(x_i, y_i)\ e(x_i, y_i) = 0,$$
$$\text{for m} = 0, 1, 2, ., \text{M}. \quad (9)$$

To solve the equation, we use the *iteratively reweighted least squares* algorithm [2]. The weighting process is governed by the scale estimate s which is updated after each fit by using the equation (6). The *Hampel redescending* function $\psi(a, b, c)$ is a three-part redescending function [2, 16] (Fig. 4). It was chosen because of the property that $\psi(x)$ = 0 for $|x| > c$, where c is a preselected cutoff value, also known as the finite rejection point, which allows rejection of outliers. Experimental results on synthetical data [4] demonstrated a high convergence speed of the robust M-estimator, and therefore the reweighting in the experiments with real data was limited to few iterations. The final solution depends on the initial fit (L2 estimate) and the scale estimate s.

4 Tracking arm silhouettes using the 2D shape model

We assume that the human head and torso remain still while the left arm is performing free movements. In the first frame, we manually segment the head and the arm segments. Next, we fit second-order curves to each region separately. These curves define the initial shape model of the arm (Fig. 2a).

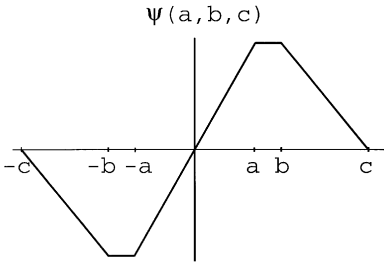


Fig. 4. Hampel redescending function $\psi(a, b, c)$

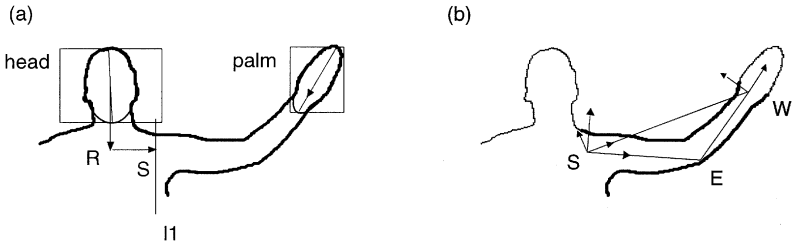


Fig. 5a,b. Initialization of the tracking procedure: **a** locating the head and the palm; **b** locating the upper and lower-arm

The subsequent image frames are processed as follows. First, the head is found automatically. Then, the root point R, the shoulder point S and the shoulder partitioning line l_1 are taken from the previous frame (Fig. 5a). The shoulder partitioning line separates the arm edge points from the rest of the silhouette. Next, the top of the arm T is located (no occlusion is allowed) and a curve is fitted to the hand data. This curve determines the location of the wrist point W. The data between the wrist point W and the shoulder point S (Fig. 5b) belongs to the upper arm and the forearm. The curves that model these segments in the previous frame are used to start the shape estimation process. At the end, a new shape model for the arm is built and the stick figure of the arm is extracted.

5 3D structure reconstruction by inversion of the kinematic model

We derived the 3D structure recovery equations by using:

1. the strong constraints of the kinematic model specified with the matrix Eqs. 1, 2 and 3, and
2. the end-points of the arm segment projections $\overline{RS}$, $\overline{SE}$ and $\overline{EW}$ extracted from the arm silhouette (Fig. 2 – left).

For a given image sequence, we assume that the lengths of the arm segments D_{cl}, D_{up} and D_{for} are known and the joint angles δ_i have to be recovered. We assume that the root point of the kinematic chain is fixed and its 3D coordinates are known. Given the 2D coordinates of the shoulder point S relative to the root point R, joint angles δ_1 and δ_2 can be calculated as follows:

$$\delta_1 = \Pi - \arccos(\frac{x_s}{D_{cl}\cos(\delta_2)}), \tag{10}$$

$$\delta_2 = \arcsin(\frac{z_s}{D_{cl}}). \tag{11}$$

From this point, the depth of the elbow point E in the root coordinate system is obtained first. Then the elbow point E is expressed in the shoulder coordinate system:

$$(x, y, z) = \mathbf{R}_y(\delta_2)^{-1} \cdot \mathbf{R}_z(\delta_1)^{-1} \cdot \vec{SE}. \tag{12}$$

Joint angles δ_3 and δ_4 are given as follows:

$$\delta_3 = \arccos(\frac{-z}{(x^2 + y^2)^{\frac{1}{2}}}), \tag{13}$$

$$\delta_4 = \arccos(\frac{(x^2 + y^2)^{\frac{1}{2}}}{D_{up}}). \tag{14}$$

From the elbow point E, the 3D coordinates of the wrist W in the root coordinate system are calculated. By expressing the wrist point W in shoulder coordinate system:

$$(x, y, z) = \mathbf{R}_x(\delta_4)^{-1} \cdot \mathbf{R}_y(\delta_3)^{-1} \cdot \mathbf{R}_y(\delta_2)^{-1} \cdot \mathbf{R}_z(\delta_1)^{-1} \cdot \vec{SW}, \tag{15}$$

joint angles δ_5 and δ_6 are given as follows:

$$\delta_5 = \arctan(\frac{-y}{x}), \tag{16}$$

$$\delta_6 = \arccos(\frac{\vec{SW}^2 - D_{up}^2 - D_{for}^2}{2D_{up}D_{for}}) - \Pi. \tag{17}$$

Due to assumed orthographic projection of the arm to the image plane, the above equations (10–17) satisfy eight solutions for the 3D position of the arm. The mirror solutions are discarded by applying the equations in the specified order.

6 Experimental results

A test sequences of monocular images were taken with a CCD camera plugged to a PC-based image-processing system. The frame rate was 5 frames/s. The human subject performed free movements with the left arm in front of a black background. The images were binarized by the image-processing system and the human silhouettes were extracted. Further processing of the image sequence was done off-line using the *Mathematica* programming environment. No camera calibration was required.

6.1 Example 1: Estimating the forearm elevation – planar motion

This experiment was carried out as follows. Three markers, assumed to determine the length of the upper arm and of the forearm, were attached to the skin of the arm. The upper arm was held still (Fig. 6), while the forearm rotated parallel to the image plane. The goal of the experiment was to compare the results for the forearm elevation calculated by using marker positions with those obtained by modeling the forearm edge points with second-order curves using robust M-estimation. The axis of the second-order curve is assumed to give the forearm elevation φ. The dimensions of the window including the data points which belong to the forearm were determined manually.

The results for the estimated elliptical shape parameters showed that the elevation parameter φ converges rapidly in

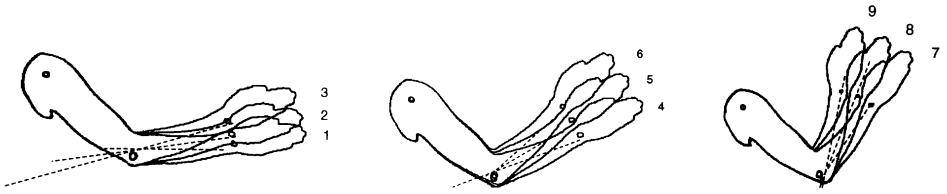


Fig. 6. Forearm elevation φ obtained by robust M-estimation

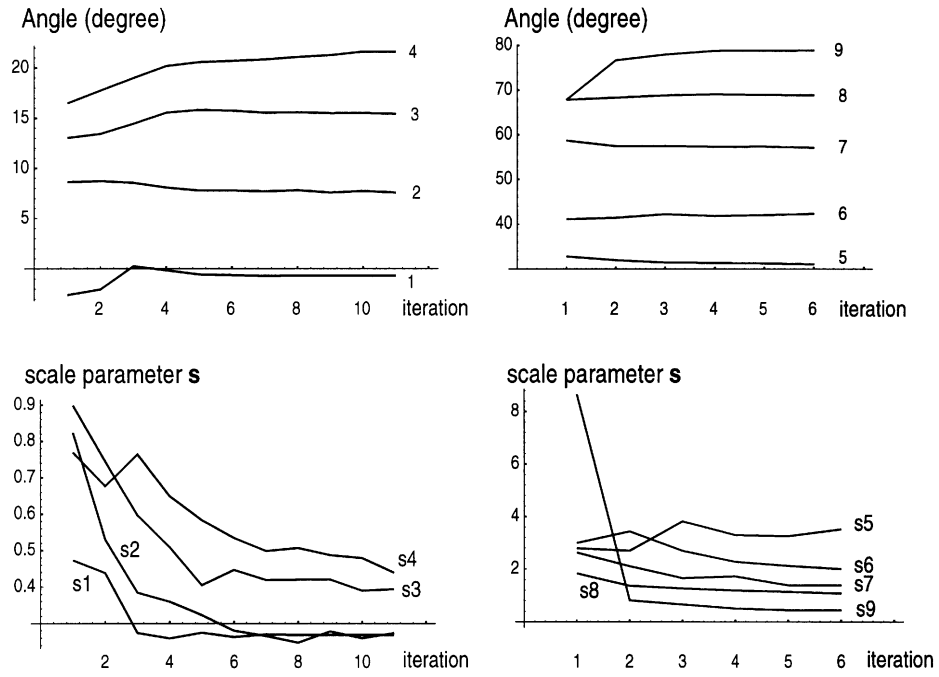


Fig. 7. Estimated elevation parameter φ and scale parameter s for the nine hand postures/frames shown in Fig. 6

Table 1. Comparison of forearm elevation φ in the frontal plane recovered by using markers and robust M-estimation

Frame	Markers	Robust estimate
1	7.0^0	$-\ 0.7^0$
2	12.2^0	7.7^0
3	20.4^0	15.7^0
4	27.0^0	20.7^0
5	36.6^0	31.1^0
6	45.4^0	42.4^0
7	53.6^0	57.1^0
8	64.4^0	68.9^0
9	75.7^0	78.9^0

the iteration process (Fig. 7). Hence, just a few iterations are enough to obtain a stable result for this parameter. Comparison of results obtained by using markers and the robust M-estimation (Table 1) correspond quite well. The difference can be explained with the displacement of the markers caused by the elasticity of the skin, as well as with the nature of the modeling process.

6.2 Example 2: accuracy analysis of 3D reconstruction for a four-d.o.f. model

In this example we used a four-d.o.f. model of the arm (three in the glenohumeral joint and one in the elbow joint). The aim was to study the accuracy of 3D reconstruction when the 2D projection of the elbow joint point is determined up to a rectangular region [6]. The robust M-estimation procedure was limited to the same number of iterations as in the previous example and applied to different sizes of windows W_1 and W_2 (Fig. 8a). Therefore, the obtained results for the orientation of the segments' axes differ. Intersections of these lines define a rectangular region inside which the projection of the elbow joint point probably lies. We simulated ten subsequent image frames of the upward arm movement. For each arm position, the rectangular uncertainty region of the elbow joint point was determined (Fig. 8b). By back-projecting these uncertainty regions into the scene, all 3D points that satisfy the constraints of the upper arm lie on a spherical surface (Fig. 8c). The depth uncertainty intervals for the position of the wrist joint are graphically presented in Fig. 8d.

6.3 Example 3: 3D motion reconstruction

The aim of this experiment was to track the 3D motion of the arm with a six-d.o.f. model of the arm. Frames (from 2 to 7) were selected from a longer sequence of 12 images (used in experiments in [5]). In this way, we simulated large displacements between frames. The actual lengths of the arm segments D_{cl}, D_{up} and D_{for} were measured beforehand on the subject and we thus calibrated the image lengths which were expressed in pixels. In the first frame, the shape model was initiated by manually mapping the rectangular regions of image data to each arm segment. As a result, the initial values for the 3D joint positions were calculated. From this

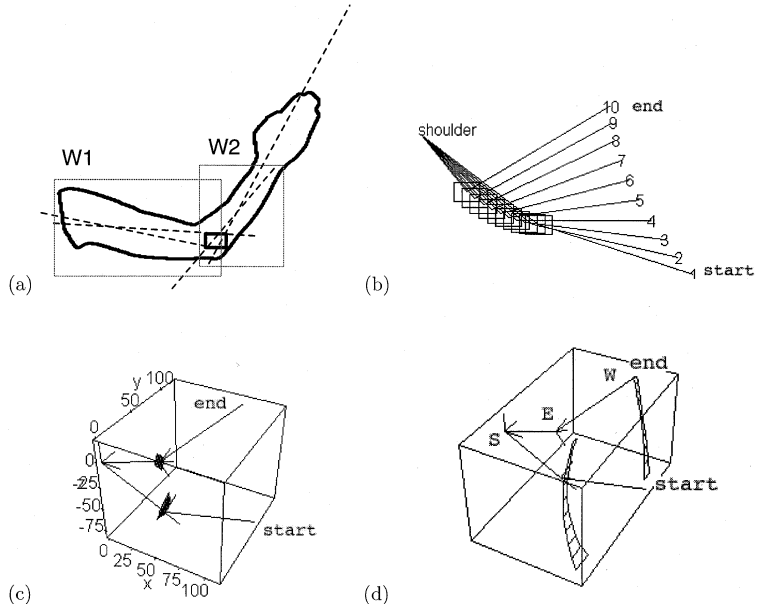


Fig. 8a–d. 3D reconstruction of the arm motion with uncertainties: **a** elbow joint position determined up to a rectangular region; **b** 10 different projections of the arm; **c** reconstructed 3D coordinates for the start and the end position of the elbow showing the uncertainty of the position; **d** two mirror 3D reconstruction solutions showing the depth uncertainty of the wrist joint point position

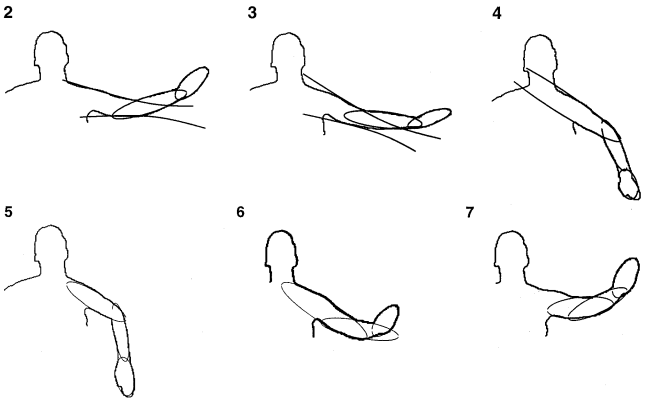


Fig. 9. Results of fitting second-order curves to the image data

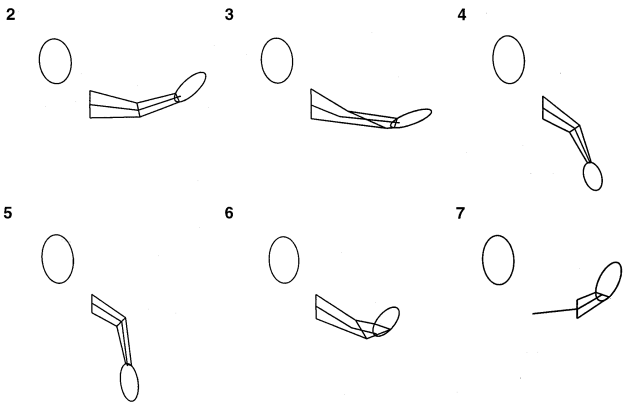


Fig. 10. Extracted models for the sequence of moves shown in Fig. 9

point on, the tracking was automatic. The results of fitting second-order curves to arm silhouettes are shown in Fig. 9. Figure 10 shows the extracted shape models. By backprojecting the extracted stick figures (Fig. 11a) into the scene, the 3D positions of the shoulder, the elbow and the wrist joint and values of the joint angles were calculated. In Fig. 11b the reconstructed trajectories of the joints are shown. From the values of the joint angles (Fig. 12), assuming that the motion is smooth, the next position of the arm in space and the image features can be predicted.

The above example shows one of the best results that we obtained. The estimated 2D shape parameters for the hand and forearm are quite stable to different orientations. However, the 2D shape parameters for the upper arm change quite a bit from frame to frame (first the upper arm is modeled with a parabola, then with an ellipse). This is influenced by the nonrigidity of the upper arm segment. As a consequence, the elbow joint point E (for example, in frame 3) is not accurately determined. In frame 6, the projection of the forearm contains a critical amount of edge points. Because there is no control over the estimation of the model parameters, the influence of outliers (edge points from the upper arm) in this case prevail.

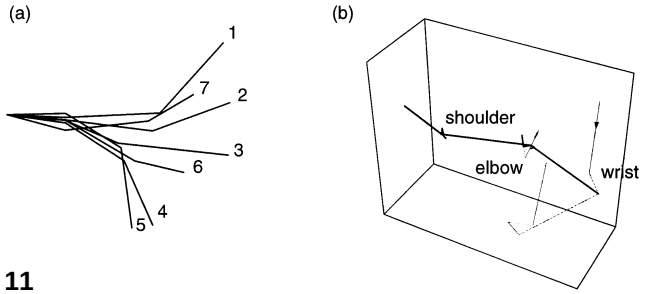
7 Discussion

The outlined approach to 3D arm motion reconstruction has the following advantages:

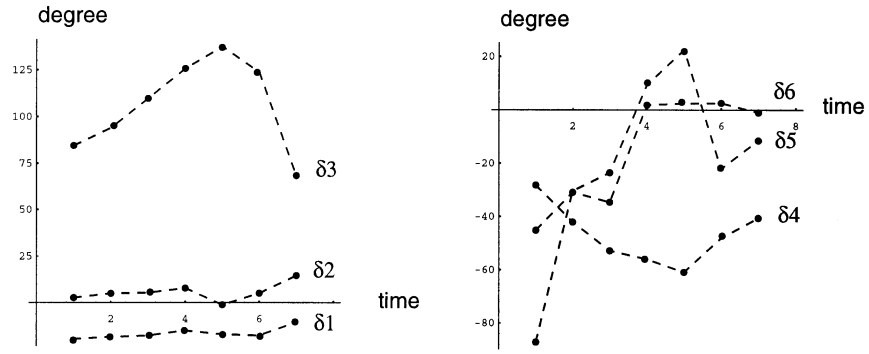
1. It is noninvasive. The subject does not need to wear markers or any other equipment.
2. It is based on a single camera. This reduces the hardware and computational costs. Also the calibration is simple.
3. Robust M-estimation is used for modeling, tracking and segmentation of 2D image data.

However, the approach does not address the following issues:

1. In the modeling process, each 2D curve is fitted separately. Better results could be obtained if we could predict/constrain the regions where 2D joints lie using multi-frame analysis.
2. It is assumed that the 2D location of the joint points can be extracted from each frame. The approach does not deal with the case where the arm parts are occluded. This drawback can be overcome by choosing a compromise solution for a joint position where it is not found and by correcting it in later frames.



11



12

Fig. 11a. Extracted stick figures of the arm, **b** Reconstructed 3D trajectories of the shoulder, elbow and wrist joint points

Fig. 12. Time-varying sequences of joint angle values

3. The 3D postures are reconstructed in a bottom-up fashion by backprojecting the extracted 2D stick figures separately in each frame. The full power of the kinematic model is thus not utilized. The kinematic model could be used in a top-down manner to predict the location of the image features by means of a Kalman filter. By interleaving the process of modeling 2D data and 3D reconstruction, better results can be expected.

8 Conclusions

We have developed a method for tracking and reconstructing 3D human arm movements from a monocular image sequence. The method is based on a six-d.o.f. kinematic model of the arm and a 2D shape model of the arm built out of second-order curves in each image frame. The novelty of the method is in tracking the arm in 2D without reconstructing the 3D surface of the arm. Our method advances the state-of-the-art techniques for tracking the human arm because the shoulder complex is modeled with 5 d.o.f. instead of 3 d.o.f. [21, 7, 8]. Our primary goal was to show that the kinematic model we used offers enough constraints for 3D reconstruction from a sequence of 2D frames. As a result, the 3D structure is solved in a computationally inexpensive way. The mathematics are simplified by assuming orthographic projection. We did not address the problem of occlusion. Before further development of the system the study of the accuracy of reconstruction using the comparison with a 3D tracking system is foreseen.

Experiments with real images of the arm performing free movements have been shown. The camera geometry was not involved in 3D reconstruction. Since we did not address the background extraction problem, uniform black background was used. To increase the applicability of the approach to more general scenes, a segmentation technique of gray-level images (for example blobs) should be employed. Another direction of improvement could be the use of multi-frame analysis to improve the stability of reconstruction. Possible applications are foreseen in 2D or 3D analysis of human motion patterns.

Acknowledgments. This research was supported by the Ministry of Science and Technology of the Republic of Slovenia (Projects J2-6187, J2-8829). The authors thank Walter Kropatsch from Vienna University of Technology in Austria for fruitful discussions of the topic. The authors are grateful to reviewers, whose comments helped to improve the quality of the manuscript.

References

1. Azarbayejani A, Pentland A (1996) Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features. In: Proceedings 13th International Conference on Pattern Recognition, Vol. C. IEEE Computer Society Press, Vienna, Austria, pp 627–632
2. Besl PJ, Birch JB, Watson LT (1988) Robust window operators. In: Proceedings 2nd International Conference on Computer Vision. IEEE Computer Society Press, Tampa, FL, pp 591–600
3. Ferrigno G, Pedotti A (1985) ELITE: A digital dedicated hardware system for movement analysis via real-time TV signal processing. IEEE Trans Biomed Eng 32:943–949
4. Filova V, Solina F, Lenarčič J (1994) Modeling 2D image data by robust statistics. In: Yüksel Ö (ed) Proceedings 7th Mediterranean Electrotechnical Conference. IEEE Press, Antalya, Turkey, pp 234–237
5. Filova V (1994) Analysis of 3D human arm motion from a monocular image sequence. M.Sc. Thesis, University of Ljubljana
6. Filova V, Solina F, Lenarčič J (1994) Image sequence analysis of 3D human arm movements. In: Pavešić N, Niemann M, Paulus D, Kovačič S (eds) Proceedings 2nd German-Slovenian Workshop, IEEE Slovenia Section, Ljubljana, Slovenia, pp 145–158
7. Gavrila DM, Davis LS (1995) 3-D model-based tracking of human upper body movement: a multi-view approach. In: Proceedings IEEE International Symposium on Computer Vision. IEEE Computer Society Press, Coral Gables, FL, pp 253–258

8. Goncalves L, Bernardo E, Ursella E, Perona P (1995) Monocular tracking of the human arm in 3D. In: Proceedings 5th International Conference on Computer Vision. IEEE Computer Society Press, Cambridge, MA, pp 764–770
9. Johanson G (1973) Visual perception biological motion and a model for its analysis. Perception Psychophysics 14:201–211
10. Hel-Or Y, Werman M (1994) Model Based Pose Estimation of Articulated and Constrained Objects. In: Eklundh J-O (ed) Proceedings 3rd European Conference on Computer Vision. Lecture Notes in Computer Science 800, Vol 1. Berlin Heidelberg New York, Springer, pp 262–273
11. Kakadiaris IA, Metaxas D, Bajcsy R (1994) Active part-decomposition, shape and motion estimation of articulated objects: A physics-based approach. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Press, Seattle, WA, pp 980–984
12. Kanatani K (1989) 3D Euclidean versus 2D non-Euclidean: two approaches to 3D recovery from images. IEEE Trans Pattern Anal Mach Intell 11(3):329–32
13. Kuch JJ, Huang TS (1995) Vision based hand modeling and tracking for virtual teleconferencing and telecollaboration. In: Proceedings 5th International Conference on Computer Vision. IEEE Computer Society Press, Cambridge, pp 666–671
14. Lenarčič J, Umek A (1993) Experimental evaluation of human arm kinematics. In: Chatila R, Hirzinger G (eds) Experimental Robotics II. Springer, London, pp 521–530
15. Maes P, Darrell T, Blumberg B, Pentland A (1995) The ALIVE System: Wireless, Full-body Interaction with Autonomous Agents. MIT Technical Report No.257
16. Mirza MJ, Boyer KL (1993) Performance evaluation of a class of M-estimators for surface parameter estimation in noisy range data. IEEE Trans Robotics Automation 9(1):75–85
17. Ohya J, Kishino F (1994) Human posture estimation from multiple images using genetic algorithm. In: Proceedings 12th International Conference on Pattern Recognition. Vol I. IEEE Computer Society Press, Jerusalem, Israel, pp 750–753
18. O'Rourke J, Badler NI (1980) Model-based image analysis of human motion using constraint propagation. IEEE Trans Pattern Anal Mach Intell 2(6):522–536
19. Pentland A, Horowitz B (1991) Recovery of nonrigid motion and structure. IEEE Trans Pattern Anal Mach Intell 13(7):730–742
20. Rehg JM, Kanade T (1994) Visual tracking of high DOF articulated structures: An application to human hand tracking. In: Eklundh J-O (ed) Proceedings 3rd European Conference on Computer Vision. Vol II, Lecture Notes in Computer Science 801. Berlin Heidelberg New York, Springer, pp 35–46
21. Yamamoto M, Koshikawa K (1991) Human motion analysis based on a robot arm model. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Press, Maui, HI, pp 664–665

Valentina Filova received a B.Sc. degree in Electrical Engineering and M.Sc. degree in Computer Science from University of Ljubljana, Slovenia, in 1990 and 1994, respectively. Her research interests are human motion analysis for biomechanics and ethology.

Franc Solina is an associate professor of computer science at University of Ljubljana and Head of the Computer Vision Laboratory at the Faculty of Computer and Information Science. He received a B.Sc. and an M.Sc. degree in Electrical Engineering from the University of Ljubljana, Slovenia in 1979 and 1982, respectively, and a Ph.D. degree in computer science from University of Pennsylvania in 1987. His research interests include range image interpretation, 3D shape reconstruction, and computer vision applications on the Internet.

Jadran Lenarčič received B.S., M.Sc., and Ph.D. degrees from the University of Ljubljana (Slovenia), Faculty of Electrical Engineering in 1979, 1981, 1986, respectively. He has been with the J. Stefan Institute since 1979 and has been Head of the Robotics Laboratory since 1985. Currently, he is Head of the Department of Automatics, Biocybernetic and Robotics. His research interests include robotics in general and robot kinematics in particular. Recently, he has been involved in projects dealing with the motion of human extremities and other investigations that relate to biomechanics and robotics. He has published about 120 journal and conference papers and has given invited lectures at many universities in Europe, Japan, and USA. He is chairing a series of international symposia on Advances in Robot Kinematics.