Journal of Neuroscience Methods 182 (2009) 97–109

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Nearly automatic motion capture system for tracking octopus arm movements in 3D space

Ido Zelman[a,*], Meirav Galun[a], Ayelet Akselrod-Ballin[a], Yoram Yekutieli[a], Binyamin Hochner[b,c], Tamar Flash[a,**]

[a] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel
[b] Department of Neurobiology, Hebrew University, Jerusalem 91904, Israel
[c] Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem 91904, Israel

ARTICLE INFO

Article history:
Received 3 September 2008
Received in revised form 26 May 2009
Accepted 27 May 2009

Keywords:
Markerless motion capture
3D reconstruction
Muscular hydrostat
Motion analysis
Octopus
Non-rigid motion
Segmentation

ABSTRACT

Tracking animal movements in 3D space is an essential part of many biomechanical studies. The most popular technique for human motion capture uses markers placed on the skin which are tracked by a dedicated system. However, this technique may be inadequate for tracking animal movements, especially when it is impossible to attach markers to the animal's body either because of its size or shape or because of the environment in which the animal performs its movements. Attaching markers to an animal's body may also alter its behavior. Here we present a nearly automatic markerless motion capture system that overcomes these problems and successfully tracks octopus arm movements in 3D space. The system is based on three successive tracking and processing stages. The first stage uses a recently presented segmentation algorithm to detect the movement in a pair of video sequences recorded by two calibrated cameras. In the second stage, the results of the first stage are processed to produce 2D skeletal representations of the moving arm. Finally, the 2D skeletons are used to reconstruct the octopus arm movement as a sequence of 3D curves varying in time. Motion tracking, segmentation and reconstruction are especially difficult problems in the case of octopus arm movements because of the deformable, non-rigid structure of the octopus arm and the underwater environment in which it moves. Our successful results suggest that the motion-tracking system presented here may be used for tracking other elongated objects.

1. Introduction

Various research domains require tracking objects and movements in space, e.g., in the analysis of motor control strategies (Borghese et al., 1996; Cutting et al., 1978), analysis of movement disorders (Legrand et al., 1998), imitation learning in robotics (Ilg et al., 2003), computer graphics (Ilg and Giese, 2002), and in vision applications, such as surveillance (Stauffer et al., 2000), driver assistance systems (Avidan, 2001) and human–computer interactions (Bobick et al., 2000). These studies use a wide range of motion-tracking techniques.

1.1. Tracking techniques

The most common automated technique for capturing human movements is based on a computerized system which tracks in real-time a set of markers that are attached at different points on the skin. This technique derives from the passive tracking of light points attached to the body (Johansson, 1973), which has evolved into a general point-light based technique for capturing biological motion (Thornton, 2006). This method generally provides efficient and accurate tracking of objects moving in 3D space. However, it usually involves expensive equipment and may be inadequate for some objects or environments. In particular, this method is inadequate for tracking movements of animals which resist having markers attached to their body or behave unnaturally with the markers attached.

Also available are markerless motion capture techniques which receive video sequences as input. Kehl and Van Gool (2006) have presented a model-based approach, which integrates multiple image cues such as edges, color information and volumetric reconstruction to fit an articulated model of the human body. Mamania et

* Corresponding author at: Faculty of Mathematics and Computer Science, Weizmann Institute of Science, POB 26, Rehovot 76100, Israel. Tel.: +972 8 9343733; fax: +972 8 9342945.
** Corresponding author at: Faculty of Mathematics and Computer Science, Weizmann Institute of Science, POB 26, Rehovot 76100, Israel.
E-mail addresses: ido.zelman@weizmann.ac.il (I. Zelman), tamar.flash@weizmann.ac.il (T. Flash).

0165-0270/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2009.05.022

al. (2004) developed a method for determining the 3D spatial locations of joints of a human body. Domain specific knowledge tracks major joints from a monocular video sequence, then various physical and motion constraints regarding the human body are used to construct a set of feasible 3D poses. Another method uses visual hull reconstruction and an *a priori* model of the subject (Corazza et al., 2006). The visual hull of an object is the locally convex approximation of the volume occupied by an object and is constructed by the projection of the object's silhouette from each of the camera planes back to the 3D volume. Wagg and Nixon (2004) developed a system for the analysis of walking human subjects in video data by extending the greedy snake algorithm (Williams and Shah, 1992) to include temporal constraints and occlusion modeling. This enabled detection of the deformable contour of humans even in cluttered environments. Chu et al. (2003) have achieved model-free markerless motion capture of human movements by extracting skeleton curves for human volumes captured from multiple calibrated cameras, then extracting human kinematics by focusing on skeleton point features.

These methods aim at capturing human motion using either a model-based or feature points approach to the problem of tracking joints and are based on the articulated rigid structure of the model of the human body. An exception is the analysis of video sequences based on the contours (the boundary curve) of the moving object. However, then only the object contour in a 2D sequence is detected, and the positions of key elements in 3D space are not provided. Fry et al. (2000) described a unique method to track free-flying insects in 3D space which has led to successful tracking of the behavior of flies and bees. Their approach tackles the challenge of acquiring data from small and fast-moving animals, such as insects in flight. However, the problems with our setting are of a different nature. As far as we know, this is the first model-free markerless motion capture approach that can automatically track octopus arm movements in 3D space.

Various computer vision techniques have also been used in different tracking problems. Avidan (2005) has considered tracking as a binary classification problem, in which trained classifiers label new pixels as belonging to an object or to the background. Wang et al. (2004) detected movement by treating video as a space-time (3D) volume of image data and using a mean shift technique to segment the data into contiguous volumes. Boiman and Irani (2005) have addressed the problem of detecting irregularities in visual data (e.g. detecting suspicious behavior in video sequences) by using a probabilistic graphical model that identifies irregular elements unlikely to be composed of chunks of data extracted from previous visual examples. Curio and Giese (2005) have combined model-based and view-based tracking of articulated figures to track human body postures in video records. However, these methods rely on either training examples or model-based information and generally aim at solving particular problems. Moreover, they all process the data as projected on a camera plane and ignore the original 3D information.

Generally, it seems that a method for motion capture must consider some aspects in advance, e.g. are we interested in a movement on a plane or in space? Are we interested just in the general position of the moving object or also in its shape? Can we model the moving object? Can we access the movement in real-time? What kind of equipment can best be used with the object and its environment? Such questions and the difficulties they raise were very relevant for our development of an automatic tracking system for octopus arm movements.

1.2. Octopus movement

Our group is conducting a large-scale research project investigating octopus motor control, focusing on octopus arm actions and investigating kinematic, biomechanical and neural aspects of movement. The octopus uses its arms for various tasks such as locomotion, food gathering, hunting and sophisticated object manipulation (Wells and Wells, 1957; Fiorito et al., 1990; Mather, 1998; Sinn et al., 2001). The efficient nature of the movements is mainly due to the flexible structure of the octopus arm which does not contain any rigid elements. Structural support and force transmission are achieved through the arm's musculature – the biomechanical principles governing octopus arm movements differ from those in arms with a rigid skeleton.

The analysis of motion of behaving octopuses by our group, particularly reaching movements (Gutfreund et al., 1996, 1998; Sumbre et al., 2001; Yekutieli et al., 2005a,b) and fetching movements (Sumbre et al., 2001, 2005, 2006), has already led to significant insights. For example, the bend point which is propagated along the arm during reaching movements was found to follow an invariant velocity profile, and the fetching movement was generalized using a vertebrate-like strategy in which the octopus arm is reconfigured into a stiffened *quasi* articulated structure. These movements were studied by analyzing the kinematics of specific, meaningful points along the arm which were found to be quite stereotypical. Electromyographic recordings and detailed biomechanical simulations assisted in revealing common principles which reduce the complexity associated with the motor control of these movements. However, kinematic description of specific points along the arm is insufficient for analyzing octopus arm movements in their full complexity. Our interest in general, as yet unclassified, movements require the analysis of the shape of an entire arm as it moves in 3D space.

Capturing the movements of the entire octopus arm raises difficult problems, mainly because of the deformable nature of the flexible arm which lacks support of rigid structures. Other difficulties in detecting octopus arm movements arise from the cluttered nature of the environment in which the octopus moves and reflections from the water or the glass when the octopus swims in an aquarium. Techniques using markers are generally inadequate, since octopuses behave unnaturally while trying to remove objects attached to their skin. Yekutieli et al. (2007) have recently presented a semi-automatic system capable of achieving accurate 3D description of a whole octopus arm in motion. As one of its first stages, this system requires manual digitization of the contours of the moving arm, a tiresome and time-consuming process, which becomes a bottleneck when a large number of movements are to be processed.

1.3. Objectives

The aim of the research presented here was to automate the system presented by Yekutieli et al. (2007) by replacing the time-consuming task of manual tracking with a movement segmentation algorithm (Akselrod-Ballin et al., 2006; Galun et al., 2003) and a smooth skeletal representation (Gorelick et al., 2004). We believe that such a system, combined with advanced electrophysiological and modeling techniques, will make a major contribution to the research on movement control of muscular hydrostats and can be used specifically as an efficient tool for assembling the large amount of data required for the study of octopus motor control. Moreover, the automated system described below presents a novel markerless motion capture technique that can capture movements of other elongated objects, e.g. the human arm and the elephant trunk. Our system is not considered model-based, since the majority of the techniques we use neither depend on nor receive any preliminary information about the moving object. The specific module which assumes an elongated characteristics of the moving object can be adapted to capture other movements in 3D space.

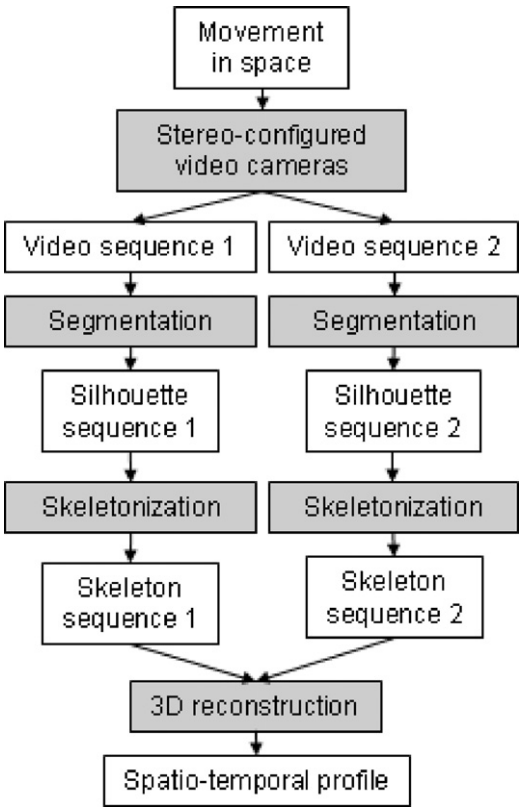

Fig. 1. Tracking motion in space. A moving object is recorded by two calibrated cameras, and each video record is processed separately by an application which tracks the motion as projected on the camera plane. A 3D reconstruction application then generates a spatio-temporal profile describing the motion in space as a function of time.

2. System framework

Our automatic motion capture system for recording and analyzing octopus behavior integrates segmentation, skeletal representation and 3D reconstruction methods as described below (Fig. 1). The input to the system is a pair of video sequences recorded by two video cameras in stereo-configuration. It is necessary to calibrate the cameras (see Section 3.3) to later reconstruct the 3D movement.

The system uses the following three main steps. First, 3D image segmentation is applied separately to each video sequence, resulting in a pair of sequences in which the segmented arm is represented by silhouettes (see Section 3.1). Using a 3D segmentation algorithm allows to analyze efficiently the whole video sequence simultaneously and not each frame separately and leads to better results. This is done by considering the length and width of the frames in the sequence as the first two dimensions and time, i.e., the duration of the video sequence as the third dimension. Then, skeletal representation is extracted for each silhouette of the arm, resulting in a pair of sequences in which the virtual backbone of the arm is prescribed by 2D curves (see Section 3.2). Finally, each pair of 2D curves that describes the arm configuration as seen by the two cameras is used to reconstruct a 3D curve, resulting in a single spatio-temporal sequence which describes the configuration of the arm in space as a function of time (see Section 3.3). Although we aim at a fully automated tracking system, two different user actions are involved in the detection process (see Section 4). However, they are applied only once during the analysis of each video sequence, independently of its length.

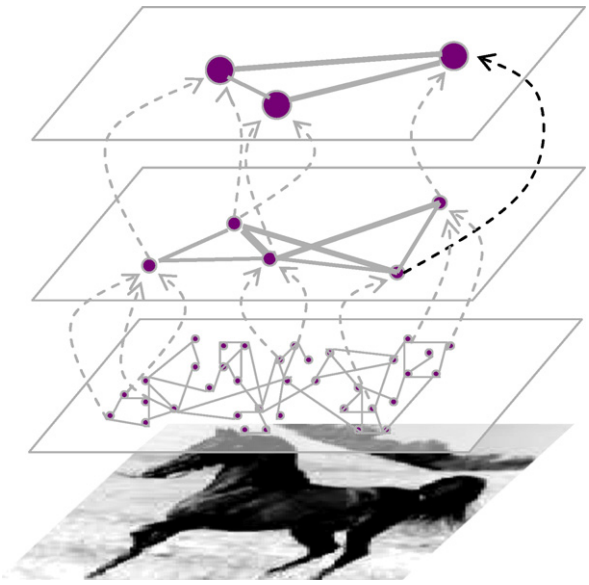

Fig. 2. A pyramidal structure represents the segmentation process in which similar parts are aggregated iteratively to form meaningful segments. See text for explanation.

3. Methods

3.1. 3D image segmentation

Here we utilize a recently presented segmentation algorithm SWA (Segmentation by Weighted Aggregation) that is capable of extracting meaningful regions of images (Sharon et al., 2001; Galun et al., 2003). The segmentation process is essentially a bottom-up aggregation framework illustrated intuitively as a pyramid structure with a dynamic number of layers for each frame in the sequence (Fig. 2). The process starts with the bottom layer that consists of the pixels of the frame and adaptively aggregates similar pixels. At first, pixels with similar intensity are merged. Then, features such as texture and shape are adaptively accumulated, affecting the aggregation process. As a result, regions with similar characteristics are aggregated into meaningful segments at the higher levels. This framework allows viewing results at very different scales of resolution (from fine to coarse), each corresponding to a different layer in the bottom-up aggregation pyramid. We now give a mathematical definition of the algorithm.

Given a (3D) video sequence, a 6-connected graph $G=(V,W)$ is constructed as follows. Each voxel i is represented by a graph node i, so $V=\{1, 2,\ldots, N\}$ where N is the number of voxels. A weight is associated with each pair of neighboring voxels i and j. The weight w_{ij} reflects the contrast between the two neighboring voxels i and j

$$w_{ij} = e^{-\alpha\left|I_i - I_j\right|}, \quad (1)$$

where I_i and I_j denote the intensities of the two neighboring voxels, and α is a positive constant. We define the saliency of a segment by applying a normalized-cut-like measure as follows. Every segment $S \subseteq V$ is associated with a state vector $u=(u_1, u_2,\ldots, u_N)$, representing the assignments of voxels to a segment S, i.e., $u_i=1$ if $i \in S$, otherwise $u_i=0$. The *saliency* Γ associated with S is defined by

$$\Gamma(S) = \frac{u^T L u}{0.5 u^T W u}, \quad (2)$$

which sums the weights along the boundaries of S divided by the internal weights. Segments which yield small values of $\Gamma(S)$ are considered salient. The matrix W includes the weights w_{ij}, and L is the Laplacian matrix of G.

Our objective is to find those partitions which are characterized by small values of Γ. To find the minimal cuts in the graph we construct a coarse version of it. This coarse version is constructed so that we can use salient segments in the coarse graph to predict salient segments in the fine graph using only local calculations. This coarsening process is recursively repeated, constructing a full pyramid of segments (Fig. 2). Each node at a certain scale represents an *aggregate* of voxels. Each *segment S*, which is a salient aggregate (i.e., $\Gamma(S)$ is low), emerges as a single node at a certain scale.

The coarsening procedure proceeds recursively as follows. Starting from the given graph $G^{[0]} = G$, we create a sequence of graphs $G^{[1]}, \ldots, G^{[k]}$ of decreasing size (Fig. 2). Similarly to the classical algebraic multigrid (AMG) setting (Brandt et al., 1982), the construction of a coarse graph from a fine one is divided into three stages: first a subset of the fine nodes is chosen to serve as the seeds of the aggregates (the latter being the nodes of the coarse graph). Then, the rules for interpolation are determined, establishing the fraction of each non-seed node belonging to each aggregate. Finally, the weights of the edges between the coarse nodes are calculated.

This segmentation algorithm can segment an object in motion by processing the frames of a video sequence simultaneously as one 3D data structure (Galun et al., 2005). It has also yielded very good results for the segmentation of 3D fMRI images which capture activity in the human brain (Akselrod-Ballin et al., 2006). Analyzing the whole sequence as one piece of data allows definition of segments not only according to their static properties, but also allows us to take advantage of the captured dynamic behavior. We have therefore chosen to process a video sequence of 2D frames simultaneously as a single 3D data structure (where the third dimension is of time and not of space). We found that this approach yields much better segmentation results than separately processing each individual frame.

3.2. Skeletal representation

Since the octopus arm shows no well-defined landmarks or features, a skeletal representation can be naturally used to model the octopus arm by curves which prescribe its virtual backbone. In Yekutieli et al. (2007) the backbone of the arm was found by a 'grass fire' algorithm in which two waves propagated from the two sides of the arm contour inwards and their loci of collision marked the middle line of the arm. The input to this method was the arm contour divided into two separate curves, the dorsal and the ventral curves. As we replaced the manual tracking by an automatic segmentation method, whose output is a silhouette of the arm and not the contour of the two sides of the arm, we had to develop a different way of extracting the skeletal representation from the silhouettes.

3.2.1. Extracting skeletal points

A common technique for extracting the skeleton of an object is the distance transform, which assigns to any point within a silhouette a value reflecting its minimal distance to the boundary contour. However, this technique may result in unsmooth skeletons that are inadequate in our case. We use an alternative method based on the solution of the Poisson equation of a given silhouette (Gorelick et al., 2004). This method comprises the following three steps:

To each internal point in a given silhouette S, we assign a value which reflects the mean time U required for a random walk from the point to hit the boundaries. This measure is computed by solving a Poisson equation of the form:

$$\Delta U(x, y) = -1, \tag{3}$$

with $(x,y) \in S$, where the Laplacian of U is defined as $\Delta U = U_{xx} + U_{yy}$, subject to the boundary condition $U(x,y) = 0$ at the bounding contour ∂S. This gives smooth level sets which can be seen as topographic lines of the silhouette (Fig. 6).

The skeletal structure is derived using U to calculate the curvature of the level set passing trough each internal point:

$$\nabla\Psi = -\nabla\left(\frac{\nabla U}{\left\|\nabla U\right\|}\right), \tag{4}$$

such that high values Ψ mark locations where level sets are significantly curved.

To further emphasize the skeleton a scale invariant version of Ψ is defined as:

$$\tilde{\Psi} = -\frac{U\Psi}{\left\|\nabla U\right\|}, \tag{5}$$

such that the small values of U near the boundaries attenuate the skeletal measure and the small values of $||\nabla U||$ near ridge locations emphasize the skeletal measure. The skeletal points are chosen as those whose value exceeds a predefined threshold (Fig. 7 presents the skeleton extracted for the silhouettes presented in Fig. 5).

3.2.2. Constructing skeletal curves

Skeletal curves prescribing the arm virtual backbone are constructed by ordering relevant skeletal points from the base to the tip of the arm, such that noisy irrelevant points are filtered. We have developed an algorithm which constructs a skeletal curve by aggregating skeletal points that form a continuous smooth curve. The aggregation of points starts from two adjacent points located in the middle arm area, and is done in two directions: towards the base of the arm and towards the tip of the arm. The algorithm consists of the following steps:

We first preprocess the skeletal points in each frame:

1. Shrinking any group of adjacent points to a minimally connected stroke.
2. Cutting short noisy branches that split from junction points.
3. Initializing the aggregation process with two adjacent points.

We then aggregate skeletal points in two opposing directions (base/tip), starting from the two initialization points, by iteratively applying the following:

Given a point c as the current point, we find a set P of candidate (nearby) points, from which the next point will be chosen:

$$P = \{p : \left|\vec{p} - \vec{c}\right| < th_{\text{dist}}\} \tag{6}$$

where th_{dist} is a predefined distance threshold. We use two measures to choose the best point from the set P: the angle θ_{past} from point c to the preceding point (irrelevant for the initialization point) and a weighted average of angles between point c and the candidate points $p_i \in P$. Then we calculate an expected angle $\tilde{\theta}$ from point c to the proceeding point as the average of the two measures:

$$\tilde{\theta} = 0.5\theta_{\text{past}} + 0.5\frac{\sum_{i=1}^{n} d_{i''}^{-1}\theta_i}{\sum_{i=1}^{n} d_{i''}^{-1}}, \tag{7}$$

where d_i is the distance from point c to point p_i, used as the weight of the angle θ_i between them. Finally, a dissimilarity score s_i is assigned to each point p_i by considering both its distance to point c (d_i), and the difference between the expected angle ($\tilde{\theta}$) to the actual angle between the points (θ_i):

$$s_i = w_d d_i + w_\theta \left|\theta_i - \tilde{\theta}\right| \tag{8}$$

where w_d and w_θ are weights (we use the standard deviations of the values $\{d_i\}_i$ and $\{\theta_i\}_i$ as weights, such that the two terms are normalized to the same scale). The point with the minimal score is chosen as the point proceeding point c. The iterative process stops when there is no further available point ($P = \phi$), and the aggregated

Fig. 3. Three representative sample frames of the octopus reaching movement. The frames correspond to the beginning (t = 50 ms), the middle (t = 500 ms) and the end (t = 950 ms) of the movement as recorded by one of two calibrated cameras. The sequence is shown after conversion to grayscales for processing (original color can be seen in Fig. 9).

points are ordered from the base to the tip of the arm forming the arm's skeleton.

3.3. 3D reconstruction

A single 3D curve is reconstructed from the pair of 2D curves, which are each the projection of the 3D curve on a different plane. This reconstruction is achieved by a recently developed 3D reconstruction application that demonstrated successful results (Yekutieli et al., 2007). The reconstruction process is based on knowledge of the projection properties for each camera, determined in advance by camera calibration. A known calibration object is recorded by each of the cameras, such that enough points are clearly visible and detectable. Marking the position of corresponding points on each of the views allows estimation of the cameras' parameters, such that the points fit the known 3D geometry of the object. The result is a projection matrix M for each of the two cameras, which relates any 3D point q to its projection p on the camera

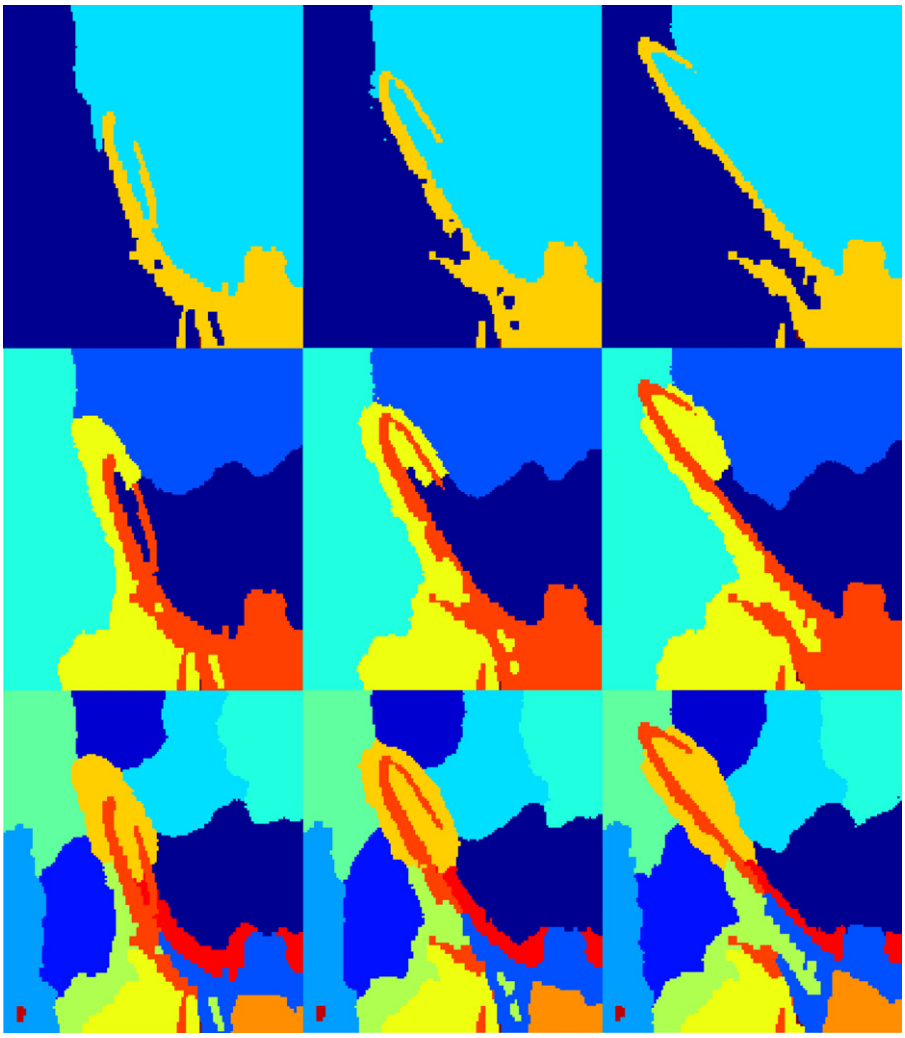

Fig. 4. 3D segmentation result for the reaching movement. The result is presented in three consecutive resolution scales, from coarser (upper) to finer (lower) segmentation.

Fig. 5. The binary sequence. The segmented arm is represented by a silhouette which consists of the appropriate parts detected by the segmentation algorithm.

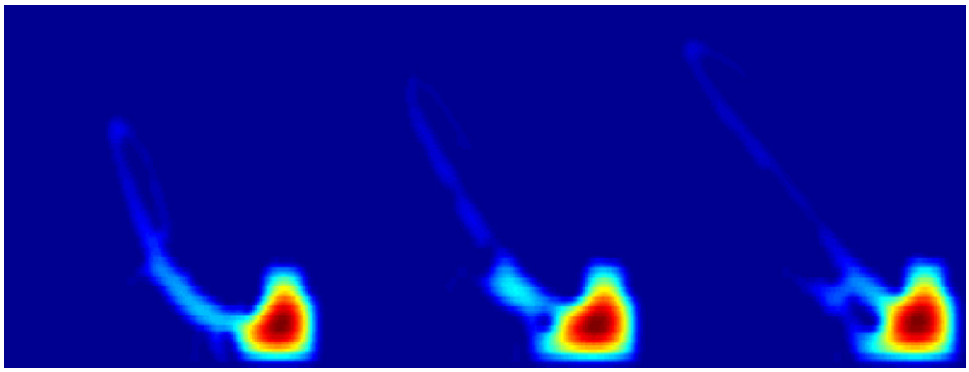

Fig. 6. The solution of a Poisson equation for the binary sequence is shown as a topographic map of the input silhouette which consists of smooth level sets. The value of each point corresponds to the mean distance from that point to the shape boundaries. These values are further processed to extract points that lie on skeletal elements of the shape.

plane expressed by $p = Mq$. Knowing the projection matrices and a pair of corresponding 2D points, one for each view, we can calculate the 3D origin of these points using the Direct Linear Transform (DLT) procedure (Abdel-Aziz and Karara, 1971; Gutfreund et al., 1996).

Each pair of skeletons is reconstructed in 3D by matching pairs of projected points that originated from the same 3D point and then reconstructing this 3D point using the DLT procedure. Corresponding points are matched using the epipolar geometry relation of the two camera planes $p_1^T F p_2 = 0$, where F is a fundamental matrix (here derived from the projection matrices), and p_1 and p_2 are a pair of the corresponding points. This relation states that for any point p_1 in one view, there is a line l_2 in the other view on which the matching point p_2 should lie, i.e., $l_2^T p_2 = 0$, and this line is expressed as $l_2^T = p_1^T F$. Therefore, the epipolar relation reduces the search for each matching pair, from a 2D search (on the whole plane) to a 1D search on that line. Here we narrow the search further by using the correct intersection between the epipolar line and the skeletal curve as the desired match. The output of this reconstruction stage is a 3D middle line curve of the octopus arm. Further details and technical descriptions of these procedures are given in Yekutieli et al. (2007).

4. Results

Octopus arm movements were video recorded while an octopus was kept in a large glass water tank. For details on animals, experimental setup and recording sessions see Gutfreund et al. (1996) and Yekutieli et al. (2007). Here we present the results of applying each of the steps described above to the video recordings of reaching movement by the octopus arm. The two input sequences are each of 1 s duration and contain 20 RGB frames, in which the octopus extends its arm toward a target. For simplicity, we present

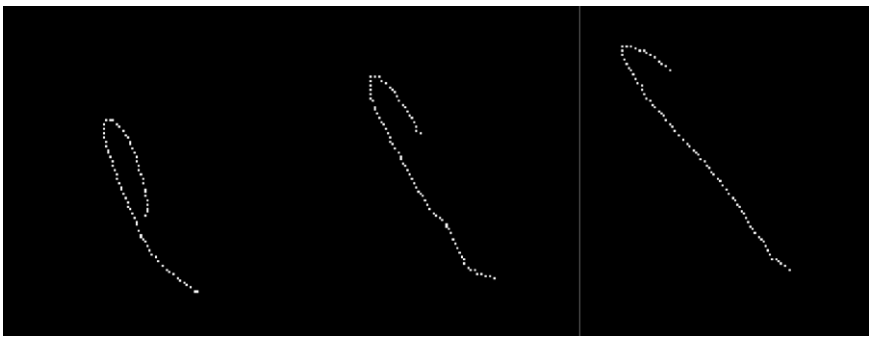

Fig. 7. A skeletal representation of the movement as projected on a camera plane. Skeletal points found by solving the Poisson equation of the segmented silhouettes are filtered and ordered, forming a continuous and relatively smooth curve for each arm configuration. These 2D curves prescribe the virtual backbone of the octopus arm during the movement.

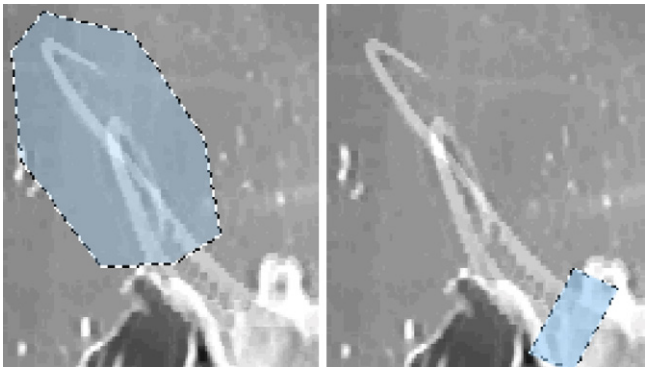

Fig. 8. The figure shows a manual procedure required by the user (once for each video sequence). In order to construct the curves prescribing the virtual backbone of the arm, the user is asked to use polygons to mark two informative domains corresponding to the areas in which the distal part (left) and the proximal part (right) of the arm moves. The algorithm uses these domains to sort skeletal points into distal and proximal groups.

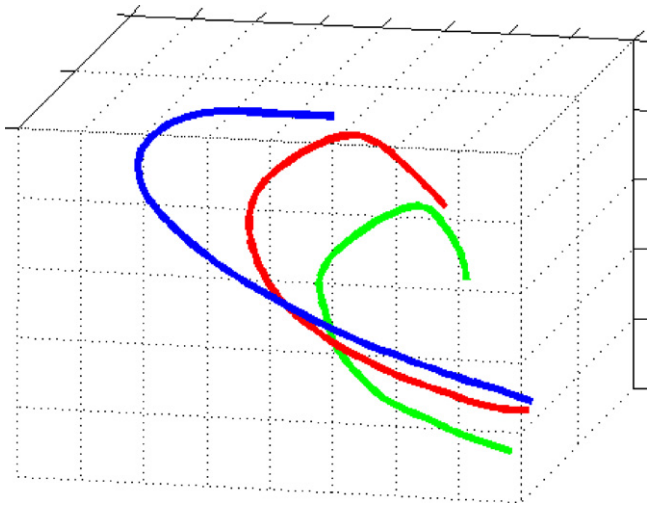

Fig. 10. 3D reconstruction of the octopus arm backbone. The presented 3D curves correspond to those in Fig. 9. Each was reconstructed from a pair of 2D curves which prescribe the same arm configuration from different angles. This final result is essentially a spatio-temporal profile.

the results of three time instances corresponding to the beginning ($t = 50$ ms), middle ($t = 500$ ms) and end ($t = 950$ ms) of the movement; the segmentation and skeletal representation results are presented for only one sequence.

The beginning, middle and end of the reaching movement sequence are shown in grayscale in Fig. 3. Conversion to grayscale allows easy improvement of poor quality video sequences using Matlab routines. However, the main reason for the conversion is that the segmentation algorithm cannot process the three RGB channels simultaneously.

Segmentation results for these frames are presented in Fig. 4. The three different resolution scales in the figure are taken from the higher part of the pyramid structure describing the process by which meaningful segments emerge through aggregating smaller segments with the same structural characteristic (see Section 3.1).

The figure shows that segments at the coarser level (higher row) are further segmented in the finer resolution scales (lower rows). The resolution scale in which the octopus arm (or any other object) is optimally segmented is not fixed and may take different values in different sequences. Therefore, at this point in the process, the user is asked to mark the optimal scale for those segments which are entirely within the movement. Generally, the optimal scale is such that the movement comprises as few segments as possible (but not necessarily a single segment). This user action must be done only once per video sequence and not for each frame.

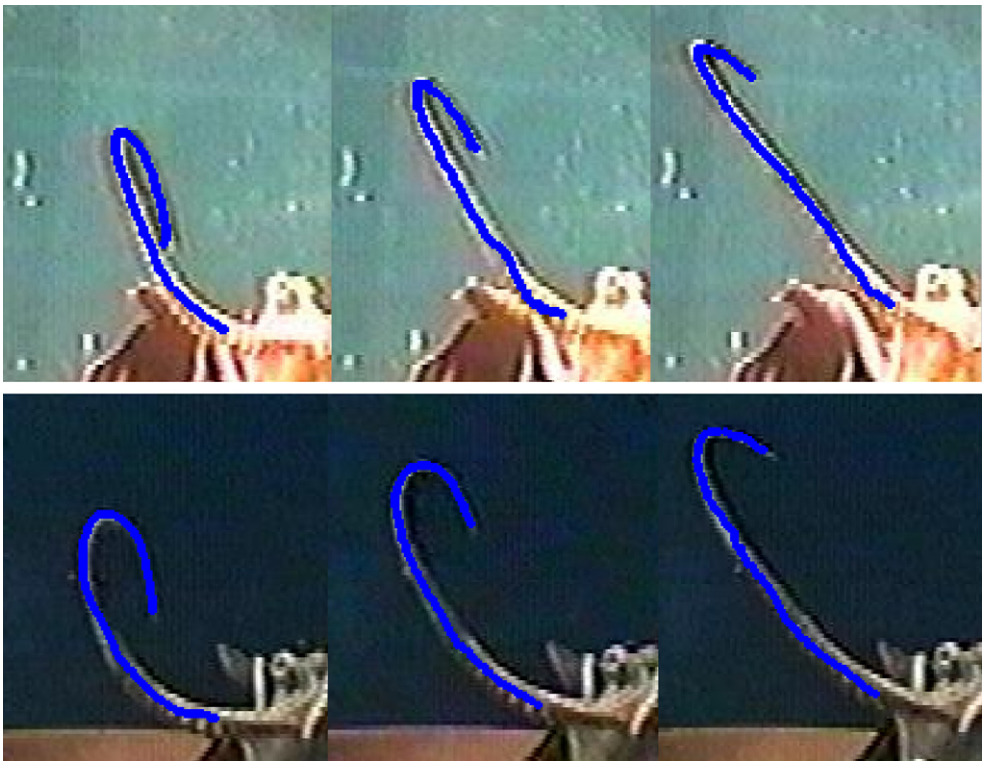

Fig. 9. Tracking results for an octopus arm reaching movement as recorded by two calibrated video cameras (a) and (b). The tracking process was applied separately for each of the two video sequences. The detected curves are processed further to reconstruct the final description of the arm movement in 3D space (see Fig. 10).

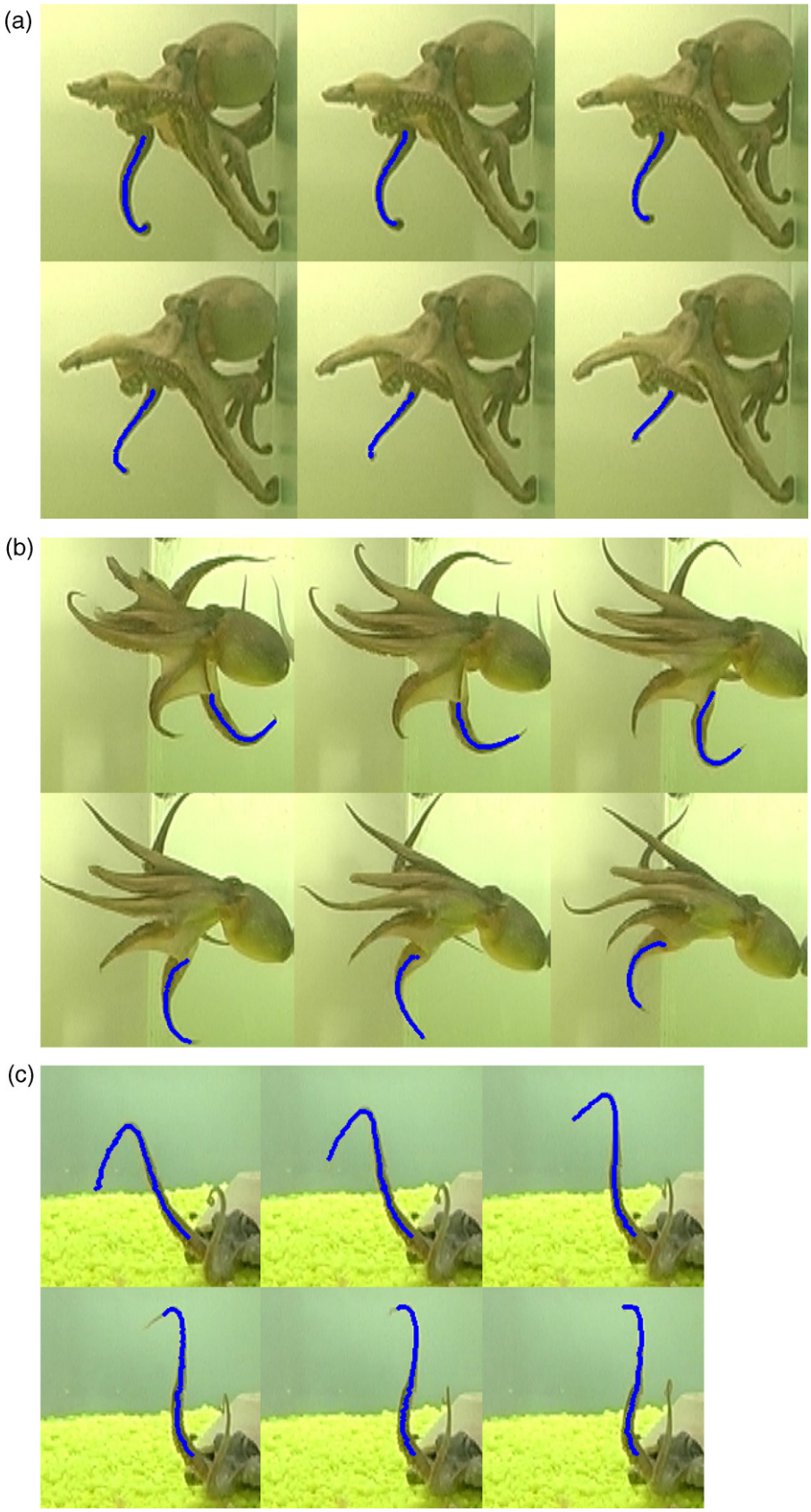

Fig. 11. Detection results for octopus arm movements not yet classified into types. The octopus performed these arm movements before moving its body (a), during a movement of the body (b) or while stationary (c–d). The movements were recorded in a setting in which unnecessary objects were removed and phosphorescent yellow stones in the background gave a relatively good contrast to the texture of the octopus skin.

(d)

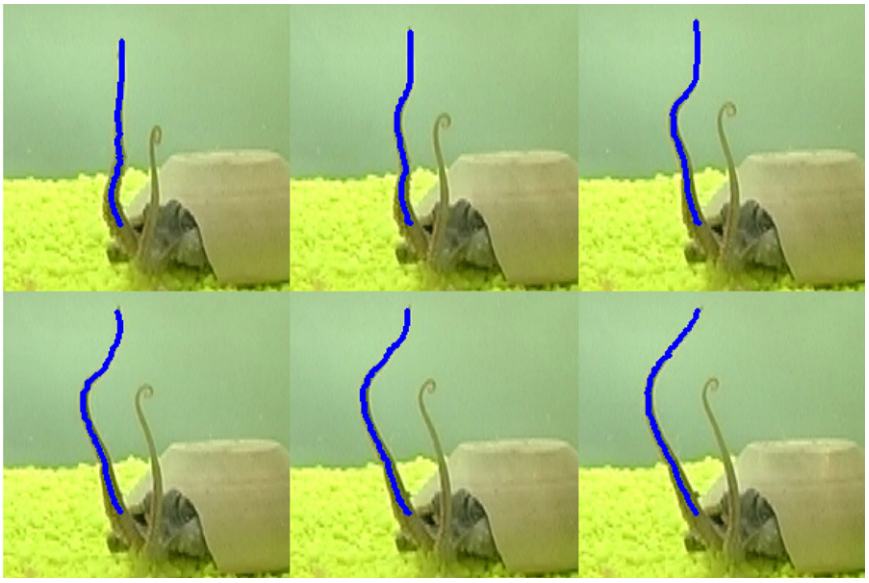

Fig. 11. (*Continued*).

Segmentation results in the generation of a sequence of silhouettes, with each silhouette standing for the octopus arm in the corresponding frame. This process is straightforward and is achieved simply by generating a binary sequence of the same size as the video sequence which is initialized by 0's, and assigning 1's for the segments that have been just marked by the user (Fig. 5).

In the next step we extract skeletal representation for each silhouette (see Section 3.2) by evaluating the average distance of internal points to the silhouette boundary (Fig. 6) and extracting skeletal point. These points are then filtered and ordered from the base to the tip of the octopus arm. The process results in a single continuous and smooth curve which prescribes the skeleton of the silhouette (Fig. 7).

The skeletal points are ordered by an aggregation process (see Section 3.2) which must be initialized by two adjacent points in each frame. This requires a user intervention. A few frames from the video sequence of the movement are presented embedded on a single frame and the user must mark with a polygon the area in which the distal part of the arm moves (Fig. 8). Then, two adjacent points for each frame of skeletal points are initialized, such that one point lies inside the polygon and the second one lies outside the polygon. As in the previous user action, this marking must be carried out only once per video sequence and not for each frame.

We have no quantitative measure to assess the quality of our results and can only suggest evaluating the results qualitatively by visual inspection. We consider a result good when the skeleton smoothly prescribes the virtual backbone of the octopus arm, starting from the base and ending as close as possible to the tip. Fig. 9a shows that this condition is fulfilled. The extracted skeletons can be easily superimposed on the corresponding input frame. Fig. 9b gives the representation of the skeleton achieved for the video sequence of the same reaching movement recorded by the second camera. Finally, Fig. 10 presents the 3D reconstruction of these three pairs of 2D skeletons as curves in 3D space. The full spatio-temporal profile that models the analyzed reaching movement of the octopus arm consists of 20 curves describing this 1 s movement in 3D space.

Next we present results for as yet unclassified octopus arm movements. Only the skeletal representation is given here, since we are demonstrating the ability of our system to automatically capture the octopus arm movements in video sequences and to integrate the results with a 3D reconstruction application. Fig. 11 presents results of four different movements performed by the octopus either before or during a movement of its body or while remaining stationary. Although we cannot assign clear functions to these movements, we are interested in reconstructing them, since these movements appear complex enough to be composed of elementary movements. Studying such movements may allow us to identify motor primitives in the motor system of the octopus (Flash and Hochner, 2005).

The quality of the detection results, particularly of the segmentation, strongly depends on the quality and characteristics of the video input sequences. The movements in Fig. 11 were recorded by better equipment (two identical and synchronized video cameras with high resolution) than used for the movement in Fig. 9. We also improved the setting by clearing unnecessary objects from the background and placing phosphorescent yellow colored stones in the water tank. This image of the color was found to contrast well with the texture and color of that of octopus skin that changes to match its environment. Although a homogenous background is optimal, it is not a required condition for high quality results as shown in Fig. 11(c and d). Fig. 11(a and b) presents high quality results where the contrast between the octopus skin and the background is relatively low.

Our system can be used to track elongated objects other than the octopus arm. The characteristics of the moving object do not have to match the non-rigid and deformable characteristics of the octopus arm. Here we present automated detection results for the motion of two other elongated objects: Fig. 12 presents tracking results for a human arm performing a free movement by detecting its approximated backbone. Fig. 13 presents tracking results for an elephant trunk drawing a picture. As with the octopus arm, the detected curve prescribes the virtual backbone of the elephant trunk.

5. Discussion

We present a novel motion capture system for 3D tracking of octopus arm movements. Since the octopus has a flexible arm lacking a rigid skeleton and with no well-defined features to be tracked, the model-based methods successfully used for the analysis of human and other object movements are not useful. Marker techniques are similarly unsuitable as they would require special equipment to cope with the octopus' underwater environment and

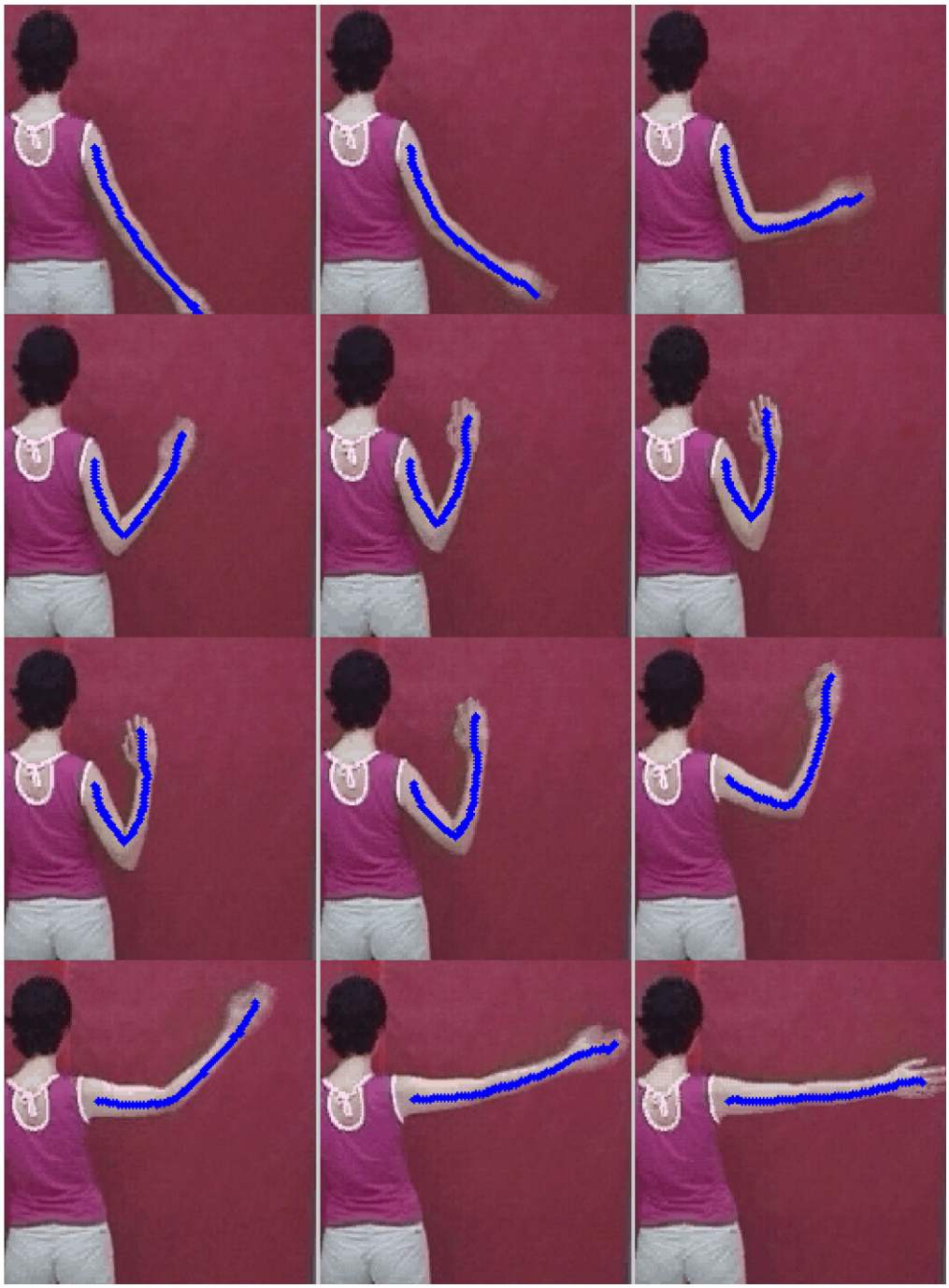

Fig. 12. Tracking results for a human arm performing a free movement (from upper left to lower right).

since the octopus does not behave naturally when markers are attached to its body or arms. Indeed, we have observed them quickly removing marker strips attached to their arms.

The motion capture system we developed to analyze octopus arm movements is thus not based on markers. Although we generally aim for a system which is not based on a model, two assumptions which refer to the shape of the analyzed object and the movement it performs are taken by the skeletalization module. The skeletalization module, which first finds the skeletal points of the extracted silhouettes regardless of their shape, assumes an elongated shape of the moving object while it sequentially orders the extracted skeletal points as a smooth 2D curve. However, since skeletalization is applied on each frame separately, assembling points in one frame does not use any of the information gathered by analyzing the former frame. The skeletalization module also assumes the ability to mark (by the user) a stationary region and a moving region for the analyzed object. This assumption essentially refers not the shape of the object but more to the characteristic of the movement it performs.

The system receives a pair of video sequences of an arm movement which has been recorded by two calibrated cameras positioned at different angles. Each video input is a projection of the movement on the 2D plane of the camera. Rather than each frame being analyzed separately, each sequence is processed as one piece of data by a segmentation algorithm that detects the silhouettes of the moving arm. Two-dimensional curves are then extracted to prescribe the virtual backbone of these silhouettes, yielding a skeletal representation of the projected movement. Finally, each pair of

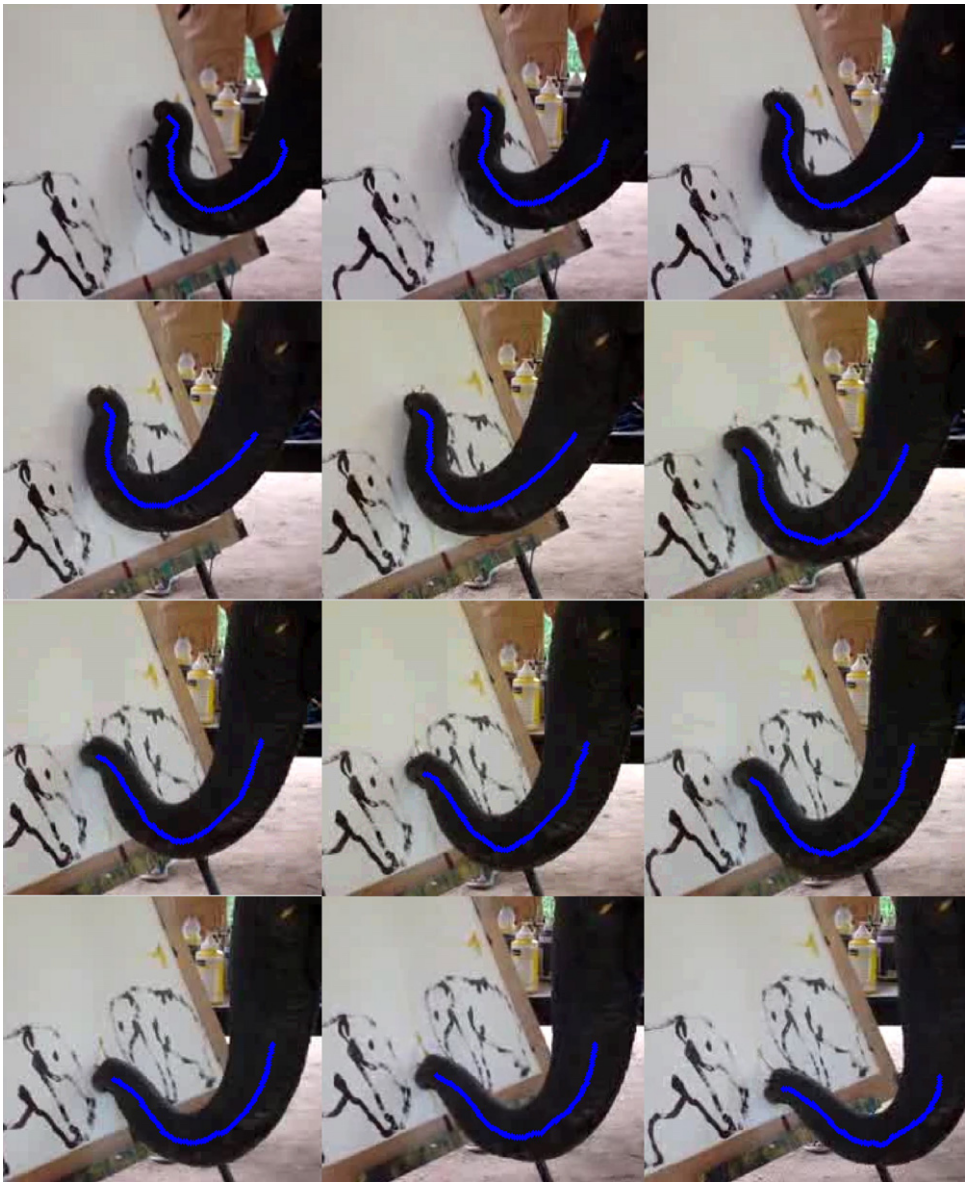

Fig. 13. Tracking results for an elephant trunk during painting (from upper left to lower right).

2D curves, describing the same arm configuration from different angles, is reconstructed into a 3D curve. This describes a temporal configuration of the arm in space, resulting in a spatio-temporal sequence of the analyzed movement. The automatic procedures used by our system allow it to efficiently analyze a large amount of data, and only two types of user actions are required once per an entire video sequence independently of its length: the first intervention occurs when the user is asked to mark the optimal scale for the results of the segmentation procedure using only a single click. In the second intervention the user is asked to use two polygons that mark two areas. One of them corresponds to the area of the stationary part of the analyzed object, e.g., the base of the octopus arm, while the second one corresponds to the moving part of the analyzed object, e.g., the distal part of octopus arm during a reaching movement (Fig. 8). We believe that these two interventions are relatively minor and hence the system can still be considered almost fully automatic, both because they are fast and easy to perform and because they refer once to the entire video sequence, regardless to the number of frames it consists of. Eventually, a 3D reconstruction of a movement requires the analysis of a pair of video sequences, therefore altogether four user interventions are required.

Our system is not restricted to the analysis of octopus arm movements, since it can process movements of both rigid and non-rigid elongated bodies. The restriction to elongated bodies is due to the procedure used by our skeletal representation method which orders skeletal points along a smooth curve as described above. Adjusting the skeletal representation module to fit other shapes allows the system to analyze movements of differently shaped bodies in 3D space. Obviously, tracking an object which can be represented by just a point is even easier, as the module which currently orders skeletal points is not needed. However, shapes like that of the octopus head, for example, have to be handled differently.

Occasionally the system may fail to automatically track the movement of the octopus arm due to problems arising in either the detection or the reconstruction stage. Automatic segmentation may fail to detect a part of the arm which is occluded by other parts of the octopus. Self-occlusion of the arm may be harder to track since the virtual backbone may no longer consist of a simple curve. The

segmentation algorithm may also fail to detect the tip of the arm, which seems to fade out in the case of a video sequence of poor quality. However, our experience shows that the level of detection of the tip is satisfying in most cases. Where the level of detection is not satisfying, we must go back and use the simpler, more robust and much more time-consuming method of manually marking the arms. Furthermore, in many octopus arm movements, the tip movements are passive, i.e., shaped mainly by the drag forces of the water. Here, our method captures the kinematics of the arm which most probably consists of the important features of the movement. The 3D configuration of the arm is generally reconstructed from the projections of the arm on the viewing planes of stereo-configured cameras. When the approximated plane on which the octopus arm lies is perpendicular to the viewing plane of one of the camera, the loss of essential information may lead to the failure of 3D reconstruction. However, the cameras are configured so as to significantly minimize these cases.

In general, the performance of our system depends on the quality of its inputs. We assume that it will be harder to track movements in an environment which is more complex due to: occlusions, light conditions, object texture and viewpoint angles. Our work in general aims for research questions, in which the sheer amount of data can compensate for small losses in either accuracy or completeness of the data. The great advantage of our automatic method is that it allows extensive analysis of movement data making a small percent of failures acceptable.

Where greater accuracy is required, the system can be improved in several ways. Incorporating model-based techniques may significantly improve its ability to cope with occluded arms and other difficult cases. We believe that a full model-based system will also be capable of dealing with reconstruction problems caused by inappropriate viewpoint angles. The segmentation algorithm, which currently processes grayscale sequences, can be further developed to simultaneously process the three colored channels of RGB sequences. Since the automatic tracking process is separately applied to each video sequence before the 3D reconstruction stage, the system can be naturally extended to a multi-camera setup (Yekutieli et al., 2007) and therefore can cope with the reconstruction problems mentioned above.

Our main interest is the analysis of different types of octopus arm movements to determine whether octopus motor control is based on the use of motor primitives. Motor primitives can be regarded as a minimized set of movements, which can be combined in many different ways to create the richness of human and animal movement repertoires and to allow learning new motor skills. Analysis for motor primitives in octopus arm movements may contribute to our understanding of how the brain handles the complexities associated with the control of hyper-redundant arms (Flash and Hochner, 2005) and may also facilitate designing control systems for hyper-redundant manipulators. Our system can significantly facilitate such analyses by allowing efficient analysis of a large number of different arm movements and modeling them. The system aims at tracking the virtual backbone which prescribes the octopus arm. It does not investigate the biomechanics of muscle contractions which results in some unique and interesting phenomena. Dealing with the complexities of octopus biomechanics would require integration of the tracking results with other methods, such as dynamic modeling. Overall the system here is a powerful tool, not only for motor control studies, but for any domain requiring motion capture and analysis.

Acknowledgments

We thank Dr. Jenny Kien for suggestions and editorial assistance and Lena Gorelick for advice and assistance. Tamar Flash is an incumbent of the Dr. Hymie Morros professorial chair.

This work was supported in part by Defense Advanced Research Projects Agency Grant N66001-03-R-8043, Israel Science Foundation Grant 580/02, and the Moross laboratory.

References

Abdel-Aziz YI, Karara HM. Direct linear transformation from comparator coordinates into object-space coordinates in close range photogrammetry. In: Proceedings of the ASP/UI Symposium on Close-Range Photogrammetry; 1971. p. 1–18.
Akselrod-Ballin A, Galun M, Gomori MJ, Filippi M, Valsasina P, Basri R, Brandt A. Integrated segmentation and classification approach applied to multiple sclerosis analysis. CVPR 2006:1122–9.
Avidan S. Support vector tracking. CVPR 2001;1:184–91.
Avidan S. Ensemble tracking. CVPR 2005;2:494–501.
Bobick AF, Intille SS, Davis JW, Baird F, Pinhanez CS, Campbell LW, Ivanov YA, Schuette A, Wilson A. The kidsRoom. Commun ACM 2000;43(3):60–1.
Boiman O, Irani M. Detecting irregularities in images and in video. ICCV 2005:462–9.
Borghese NA, Bianchi L, Lacquaniti F. Kinematic determinants of human locomotion. J Physiol 1996;494(3):863–79.
Brandt A, McCormick SF, Ruge JW. Algebraic multigrid (AMG) for automatic multigrid solution with application to geodetic computations. Fort Collins, Colorado: Inst. for Computational Studies, POB 1852; 1982.
Chu C, Jenkins O, Matari M. Towards model-free markerless motion capture. IEEE Int Conf Robot Autom 2003.
Corazza S, Mundermann L, Chaudhari AM, Demattio T, Cobelli C, Andriacchi TP. A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach. Ann Biomed Eng 2006;34(6):1019–29.
Curio C, Giese MA. Combining view-based and model-based tracking of articulated human movements. In: Proceedings of the IEEE Workshop on Motion and Video Computing (WACV/MOTION), vol. 2; 2005. p. 261–8.
Cutting JE, Proffitt DR, Kozlowski LT. A biomechanical invariant for gait perception. J Exp Psychol Hum Percept Perform 1978;4(3):357–72.
Fiorito G, Planta CV, Scotto P. Problem solving ability of *Octopus vulgaris* Lamarck (Mollusca, Cephalopoda). Behav Neural Biol 1990;53:217–30.
Flash T, Hochner B. Motor primitives in vertebrates and invertebrates. Curr Opin Neurobiol 2005;15(6):660–6.
Fry SN, Bichsel M, Muller P, Robert D. Tracking of flying insects using pan-tilt cameras. J Neurosci Methods 2000;101:59–67.
Galun M, Apartsin A, Basri R. Multiscale segmentation by combining motion and intensity cues. CVPR 2005:256–63.
Galun M, Sharon E, Basri R, Brandt A. Texture segmentation by multiscale aggregation of filter responses and shape elements. ICCV 2003:716–23.
Gorelick L, Galun M, Sharon E, Basri R, Brandt A. Shape representation and classification using the poisson equation. CVPR 2004:61–7.
Gutfreund Y, Flash T, Fiorito G, Hochner B. Patterns of arm muscle activation involved in octopus reaching movements. J Neurosci 1998;18:5976–87.
Gutfreund Y, Flash T, Yarom Y, Fiorito G, Segev I, Hochner B. Organization of octopus arm movements: a model system for studying the control of flexible arms. J Neurosci 1996;16:7297–307.
Ilg W, Giese MA. Modeling of movement sequences based on hierarchical spatial-temporal correspondence of movement primitives. Biol Motiv Comput Vision 2002:528–37.
Ilg W, Bakir GH, Franz MO, Giese MA. Hierarchical spatio-temporal morphable models for representation of complex movements for imitation learning. In: Proceedings of the 11th International Conference on Advanced Robotics, University of Coimbra; 2003. p. 453–8.
Johansson G. Visual perception of biological motion and a model for its analysis. Percept Psychophys 1973;14:201–11.
Kehl R, Van Gool LJ. Markerless tracking of complex human motions from multiple views. Comput Vision Image Underst 2006;103:190–209.
Legrand L, Marzani F, Dusserre L. A marker-free system for the analysis of movement disabilities. Medinfo 1998;9(2):1066–70.
Mamania V, Shaji A, Chandran S. Markerless motion capture from monocular videos. In: Proceedings of the Fourth Indian Conference on Computer Vision, Graphics & Image; 2004. p. 126–32.
Mather JA. How do octopuses use their arms? J Comp Psychol 1998;112:306–16.
Sharon E, Brandt A, Basri R. Segmentation and boundary detection using multiscale intensity measurements. CVPR 2001;I:469–76.
Sinn DL, Perrin NA, Mather JA, Anderson RC. Early temperamental traits in an octopus (*Octopus bimaculoides*). J Comp Psychol 2001;115(4):351–64.
Stauffer C, Eric W, Grimson L. Learning patterns of activity using real-time tracking. IEEE Trans Pattern Anal Mach Intell 2000;22(8):747–57.
Sumbre G, Fiorito G, Flash T, Hochner B. Motor control of the octopus flexible arm. Nature 2005;433:595–6.
Sumbre G, Fiorito G, Flash T, Hochner B. Octopuses use a human-like strategy to control precise point-to-point arm movements. Curr Biol 2006;16(8):767–72.
Sumbre G, Gutfreund Y, Fiorito G, Flash T, Hochner B. Control of octopus arm extension by a peripheral motor program. Science 2001;293:1845–8.
Thornton IM. Biological motion: point-light walkers and beyond. In: Knoblich, et al., editors. Human body perception from the inside out. Oxford: Oxford University Press; 2006. p. 17–20.
Wagg DK, Nixon MS. Automated markerless extraction of walking people using deformable contour models. Comput Anim Virtual Worlds 2004;15(3–4):399–406.

Wang J, Thiesson B, Xu Y, Cohen M. Image and video segmentation by anisotropic kernel mean shift. In: Proceedings of the European Conference on Computer Vision, vol. II; 2004. p. 238–49.
Wells MJ, Wells J. The function of the brain of octopus in tactile discrimination. J Exp Biol 1957;34:131–42.
Williams D, Shah M. A fast algorithm for active contours and curvature estimation. CVGIP: Image Underst 1992;55:14–26.
Yekutieli Y, Sagiv-Zohar R, Aharonov R, Engel Y, Hochner B, Flash T. Dynamic model of the octopus arm. I. Biomechanics of the octopus reaching movement. J Neurophysiol 2005a;94(2):1443–58.
Yekutieli Y, Sagiv-Zohar R, Hochner B, Flash T. Dynamic model of the octopus arm. II. Control of reaching movements. J Neurophysiol 2005b;94(2):1459–68.
Yekutieli Y, Mitelman R, Hochner B, Flash T. Analysis octopus movements using three-dimensional reconstruction. J Neurophysiol 2007;98:1775–90.