SMART3D: TRACKING OF MOVEMENTS SURVEYED BY A MULTIPLE SET OF TV-CAMERAS

N.A. Borghese

Istituto di Fisiologia dei Centri Nervosi - Consiglio Nazionale delle Ricerche,
Via Mario Bianco, 9, 20131 Milano, Italy.

INTRODUCTION

The systems for automatic movement analysis adopting passive markers offer the great advantage that do not bear any encumbrance to the subject. This is a necessary prerequisite to collect reliable quantitative data on sport gestures. These systems feed the host computer with the two dimensional target co-ordinates of the markers, ordering them always from the top to the bottom of the picture, without any relation with the corresponding body points. The two dimensional points must be assigned to their corresponding markers and their three-dimensional location reconstructed before any further processing. With these data, a model of the body during the movement can be pictured. Assigning the two-dimensional points to their corresponding three-dimensional markers is a task that greatly increases in complexity when more than two TV-cameras are adopted. The complexity also greatly increases when some of the markers are hidden for some frames due to overlapping segments or when their trajectories become very close one to the other. In some frames spurious markers may appear and they should be correctly recognised as reflexes and eliminated. Moreover when the hidden markers reappear, the system should be able to correctly classify them. We pursue here a modular sequential approach which is based on a frame by frame evaluation of the affinity of two-dimensional points with three-dimensional markers. The system that implements these procedures has been named SMART3D (System model-based for the Automatic Reconstruction of Trajectories in 3D).

METHOD

SMART-3D system works in two different domains that we will call the model domain (3D co-ordinates) and the target domain (2D co-ordinates of the markers). The markers surveyed are processed through four main hierarchical steps, frame by frame (figure 1).

For our experiments, we used a configuration of 4 TV-cameras: two of them located in front of the subject and the other two above her. The bidimensional co-ordinates of the markers have been collected by the Elite System, 100 Hz (Ferrigno et al., 1991).

First step

First of all, the prevision of the position of all the markers is computed starting from their positions in the previous frames. An autoregressivc model with three parameters is used; the parameters are computed from the data in the previous six frames.

$$x(t+1) = a1 * x(t) + a2 * x(t-1) + a3$$

where a1, a2 and a3 are computed from:

$$x(t-k) = a1 * x(t-k) + a2 * x(t-k-1) + a3 \quad k=0,1,2,3,4$$

The determinant of the coefficients matrix is compared to a threshold; if it is under threshold, the AR model is overdetermined and a straight line is the best fitting to that segment of trajectory; as a special case, it is a constant position value.

These three-dimensional points are projected on the TV-cameras target and two-dimensional regions of possible location are defined for each marker.

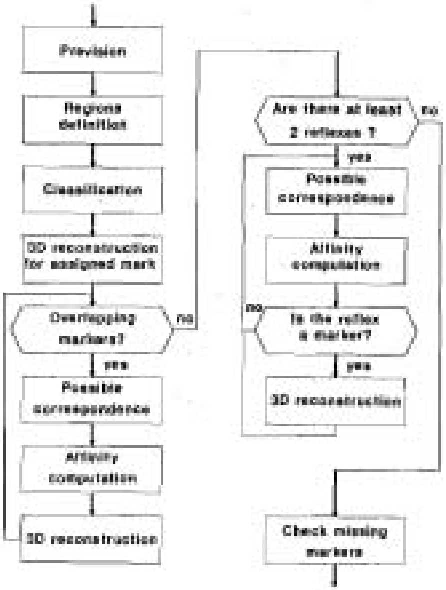

Figure 1. Flow chart of the algorithm.

A classification of the state of the frame is now carried out; a marker is labelled "assignable" if there is only that marker in one region and that region does not intersect with any other. A marker is labelled "overlapping" if the region to which it belongs to intersects with any other; a 'reflex" is a marker that has no corresponding region and a "missing" marker is a region that does not contain any marker in that frame.

Second step

If an assignable marker is present on at least two TV-cameras, its three-dimensional position can be reconstructed. The "overlapping" markers are processed in the third step.

Third step

In order to solve the correspondence problem for overlapping markers, the straight lines through perspective centres and two dimensional points not yet assigned to a three-dimensional marker are build and their distance is computed.

If this distance is less then a threshold value, the two dimensional points are candidate to be a projection of the same three-dimensional marker. This is a very hard constraint, it is a stereopsis constraint that eliminates a lot of false correspondences.

Afterwards, the three-dimensional distances between all the markers is estimated starting from their distances in the previous frames; when there are small fluctuations on the distance during the movement, the connection between two markers is supposed to be rigid, otherwise the link is considered flexible. Notice that this is a run time estimation of rigidity.

For each couple of 2D points which is candidate to be a three-dimensional marker, an affinity is computed.

$$A_i = \frac{a * \sum_{1}^{N_{ri}} \left| |P_i - P_j| - d_{ij} \right|}{N_{ri}} + \frac{b * \sum_{1}^{N_{si}} \left| |P_i - P_k| - d_{ik} \right|}{N_{si}} + c * \left| P_i - \overline{P}_i \right|$$

Ai indicates the affinity of the candidate point i with a marker not yet assigned. dij is the length estimation of segment connecting point with marker Pj, Nri is the number of connections to marker i estimated to be rigid and Nsi is the number of connections estimated flexible. Pi is the position of the marker tested. a, b and c are weighing coefficients.

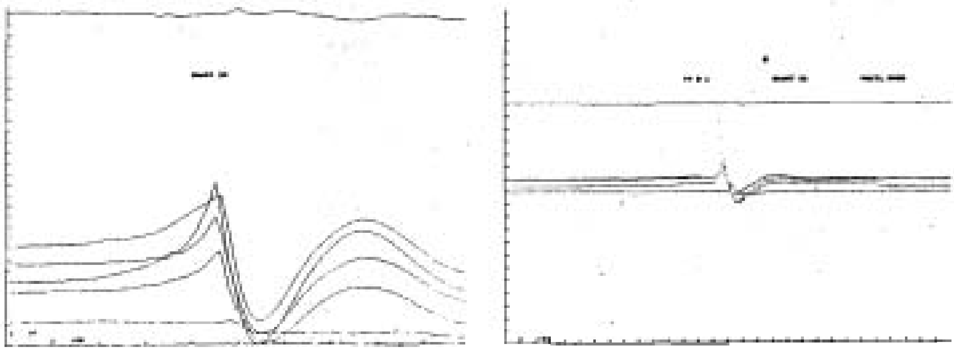

Figure 2a, 2b - A movement like playing piano in the air with five markers put on the thumb, metaphalangeal joint, index and middle finger. The vertical co-ordinate of raw data is plotted, as it was surveyed and after the 3D reconstruction.

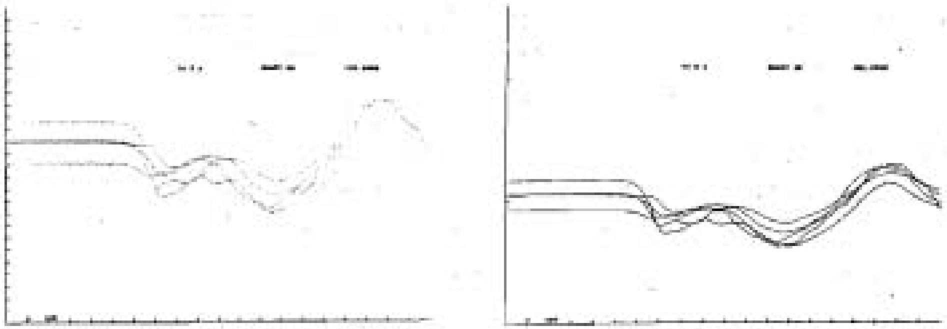

Figure 3 - The task of catching a falling ball in which shoulder, elbow, fingers and the falling ball are marked, on the left the raw data surveyed by one TV-camera, on the right the projection of the 3D reconstructed positions on the target of the same camera.

In impulsive movements the 3rd term is neglected and affinity becomes a distance test.

Fourth step

The fourth main step deals with markers that disappeared for some frames. When at least two bidimensional markers are labelled reflexes on at least two TV-cameras and there are missing markers, a congruence test is performed to see if these bidimensional points can be one of the markers momentarily disappeared. This test is based on the distribution of the distances between the markers. At the end of the tracking procedure the system is able to give a tentative structure of the moving object in terms of links connected by hinges by considering links only those segments connecting two markers whose three-dimensional distance was approximately constant.

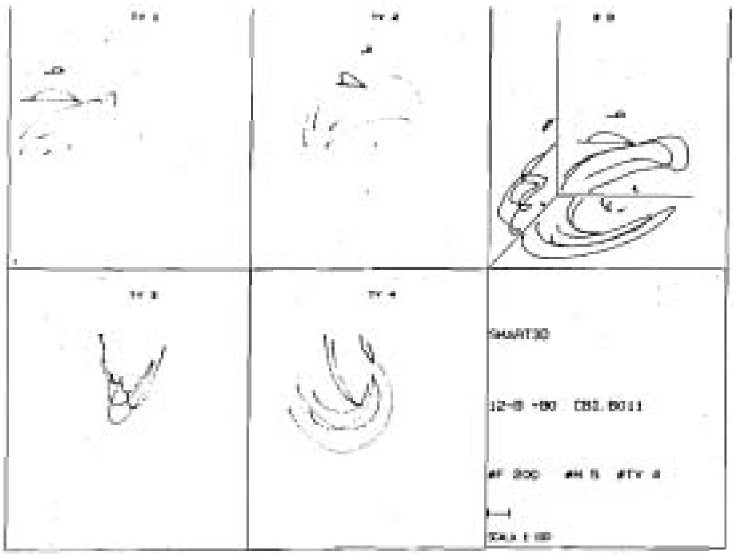

Figure 4 - A rotatory movement of the arm and the head.

CONCLUSIONS

This system has been tested in our laboratory using a set up of 4 TV-cameras working completely independently. It allows to assess all those movements for which the classical two TV-cameras set-up is insufficient. Moreover is also proved a higher reliability on tracking data from a couple of TV-cameras with respect to 2D algorithms because it relies on the 3-D distances between the markers along with the stereopsis constraint that avoid false matching of the markers.

REFERENCES

Borghese N.A. and Ferrigno G., (1990) An Algorithm for 3D automatic movement analysis by means of standard TV-cameras, IEEE Trans. on Biom. Eng., pp. 1221-1225.

Ferrigno G, Borghese N.A., Pedotti A., (1991). Pattern recognition in 3D automatic human motion analysis, special issue "Medical Imaging and Stereometry" of ISPRS J. Photogramm. and Remote Sensing.

ACKNOWLEDGEMENT: This work was supported in part by CEE Grant 3149 ESPRIT II Basic Research.