Low Density Feature Point Matching for Articulated Pose Identification

Horst Holstein and Baihua Li
Department of Computer Science
University of Wales, Aberystwyth
Aberystwyth, Ceredigion, SY23 3DB, Wales, UK
hoh@aber.ac.uk, bbl00@aber.ac.uk

Abstract

We describe a general algorithm for identifying an arbitrary pose of an articulated subject with low density feature points. The algorithm aims to establish a one-to-one correspondence between two data point-sets, one representing the model of an observed subject and the other representing the pose taken from freeform motion of the subject. We avoid common assumptions such as pose similarity or small motion with respect to the model, and assume no prior knowledge from which to infer an initial or partial correspondence between the two point-sets. The algorithm integrates local segment-based correspondences under a set of affine transformations, and a global hierarchical search strategy. Experimental results, based on synthetic pose and real-world human motion capture data demonstrate the ability of the algorithm to perform the identification task. Reliability is compromised as noisy data and limited segmental distortion are increased, but the algorithm can tolerate moderate levels. This work therefore contributes to establishing an initial correspondence in point-feature tracking for articulated motion.

1 Introduction

Identification of an articulated pose arises naturally from object tracking and recognition [1]. The articulated motion we are considering describes segment-based jointed motion, such as occurs in vertebrate biological motion. The motion of each segment can be considered as rigid or nearly rigid, but the whole motion is high-dimensionally non-rigid. Investigations of articulated movements have been a growing interest in the past decade, motivated by potential applications such as human-computer interfaces, biomedical studies, the entertainment industries and robotics.

When an articulated motion is represented by a sequence of feature points, the spatio-temporal information of the articulated motion is reduced to a sequence of moving points over time. Of the fundamental tasks in point-feature tracking systems, such as feature detection, 3D reconstruction and inter-frame tracking have been investigated extensively [5, 15, 6, 7, 14]. However identification, to know which point in an observed data corresponds to which point in its model, still remains an open problem, especially at the start or recommencement of tracking. Currently, most tracking approaches deal with in-

BMVC 2002 doi:10.5244/C.16.66

cremental pose estimation relying on manual initialization, or on an assumption of initial pose similarity to the model, or knowledge related to a specific motion.

We concentrate on the identification task to address the problem of self-initialization in 3D point-feature tracking systems. Therefore, our algorithm assumes availability of 3D feature point data, such as obtained by stereo-vision techniques or other sensors. We use a model-based point pattern matching approach.

Point pattern matching (PPM) is a fundamental yet still open problem in computer graphics, computer vision and pattern recognition [8, 11, 10, 4, 3, 18], more often restricted to rigid, affine and projective point matching. Rangarajan et al.[12] described a method of non-rigid point matching to register two contours composed of dense point sets. They approximate articulation by simultaneously determining a set of matches and piecewise-affine transformations. Song et al. [16] addressed the problem of a self-initializing tracker based on a probabilistic modeling of human walking motion from training data in images. Commercial marker-based optical motion capture (MoCap) systems [13] are currently used in medical studies, sports analysis and animation etc. They may fail in the task of identification at times, requiring manual post-processing before the data are useable in actual applications. Systems generally fall short of dealing with real-world situations, such as: i) articulated one-to-one matches in the presence of missing (due to occlusion) and noisy data (arising from measurement), ii) distortion of non-rigid segments, iii) complete independence on initial location and pose.

In this paper, we present a segment-based articulated point matching (SAPM) algorithm to automatically identify an arbitrary pose from 3D low density feature points, rather than dense point features on a curve, contour or surface [3, 18, 12]. In our experiments, a segment-based skeletal model of an observed subject is manually generated off-line using one frame of pose data that includes full feature points. Feature points are located not only on joints between articulated segments [6, 16, 14], but must be sufficient in number to indicate detailed structure and orientation, e.g. as shown in Fig.2 and 4. The observed pose data are taken from one frame of noisy feature-point pose data during the subject's freeform movements. The identification task is to find the precise one-to-one matches between the points in the subject model and the points in its observed pose data. We successfully address this problem in real-world situations with occlusion, noisy data and limited distortion in each segment. The proposed algorithm contributes to articulated pose estimation for self-initializing three dimensional point-feature tracking systems. It aims at complete independence of manual intervention for identifying unexpected and arbitrary poses, either at the beginning or on resumption of tracking.

2 Problem statement and formulation of the objective

When we use 3D sparse points as features to locate and represent an articulated motion, the geometrical structure of the articulated subject is modeled by a high-degree-freedom skeleton with a set of rigid segments between articulated joints, e.g. as shown in Fig.2 (left). To keep the articulated model general, we allow each segment of the articulated subject to undergo a set of independent affine transformations within the constraint of a jointed skeleton, and furthermore allow limited distortion in "rigid " segments. We do not impose range of motion constraints such as feasible biological motion.

More formally, we state the matching problem as: within the space $\Omega \in \{\Re^3\}$, given

are two point-sets. One is the model. This model consists of a set of identified feature points described by their 3D coordinates $p_{s,i} \in \Omega$, their labels $L_{s,i}$, and their grouping into a set of S segments: $P = \{P_s | \, P_s = \{(p_{s,i}, L_{s,i}), i = 1, ...M_s\}, s = 1, ...S\}$, each segment having M_s model points. The other is the observed set $Q = \{q_i | \, q_i \in \Omega, i = 1, 2, ...N\}$ of N unidentified points in a randomly sampled frame of the modeled subject during its movements.

The SAPM algorithm includes two steps: firstly, to establish a best local segment-based one-to-one correspondences by looking for a set of affine transformations that interpret segmental movements; and secondly, to apply a global hierarchical search strategy.

To define a set of best local correspondences between the two point-sets, we utilize two criteria: matching quality $\bar{e}$ and matching size Θ. For matching quality, we require that the mean matching error $\bar{e}_s$ of prototype segment $P_s(p_{s,i})$ and its assumed match $Q_s(q_{s,i})$ be less than the tolerable segmental-distortion ε in Eq.2, under an affine transformation $[R_s, T_s]$ (R_s for rotation and T_s for translation) defined by minimizing objective function in Eq.1.

$$[R_s, T_s] = \arg \min_{R_s, T_s} \frac{\sum_{i=1}^{M_s} r_{s,i} \times ||q_{s,i} - R_s p_{s,i} - T_s||}{\sum_{i=1}^{M_s} r_{s,i}} \tag{1}$$

$$\bar{e}_s = \frac{\sum_{i=1}^{M_s} r_{s,i} \times ||q_{s,i} - R_s p_{s,i} - T_s||}{\bar{l}_s \times \sum_{i=1}^{M_s} r_{s,i}} < \varepsilon \tag{2}$$

For the matching size condition Θ_s, we require that subset Q_s have at least $\beta \times M_s$ ($\beta \in [0, 1]$) non-null matches (Eq.3).

$$\Theta_s = \sum_{i=1}^{N_s} r_{s,i} > \beta \times M_s \tag{3}$$

When both conditions are satisfied, we regard $(P_s(p_{s,i}), Q_s(q_{s,i}))$ as the best segmental correspondence. In equations Eq.1- 3 above, the norm $||q_{s,i} - R_s p_{s,i} - T_s||$ denotes the Euclidean distance between matching point $q_{s,i}$ and its transformed model $p_{s,i}$, and $\bar{l}_s = \overline{\sum_{i,j} d(p_{s,i}, p_{s,j})}$ denotes the average segmental length. We indicate a non-null match $(q_{s,i}, p_{s,i})$ by $r_{s,i} = 1$, otherwise we set $r_{s,i}$ to 0.

Global articulated matching is achieved by integrating local segment-based matching with a hierarchical search strategy (section 3.2).

3 Algorithm

3.1 Local segment-based matching

Articulated motion maintains geometric invariance in the "rigid" segments. Local matching is therefore possible at the segment level. The algorithm generates candidate tables (CTs) of matching points to a given segments. A relaxed criterion is used since "rigid" segments are allowed some degree of distortion. The tables are then prioritized for iterative matching.

3.1.1 Candidate-tables generation

If no point has been identified for an unmatched segment P_s, we arbitrarily choose a pivot $p_{s,l} \in P_s$ and order the remaining $p_{s,i} \in P_s$ by non-decreasing distance from the pivot to define an ordered pivotal sequence for segment P_s. Then choose a match candidate q_k from the unidentified observed points $\overline{Q}_M$ for the pivot $p_{s,l}$, and generate a column of match candidates q_j for each non-pivotal element $p_{s,i}$ of the pivotal sequence, according to Eq.4, where $d(p_{s,l}, p_{s,i})$ denotes Euclidean distance.

$$\frac{|d(p_{s,l}, p_{s,i}) - d(q_k, q_j)|}{d(p_{s,l}, p_{s,i})} < 2\varepsilon; \;\; q_j \in \overline{Q}_M \tag{4}$$

The resulting array is a candidate table CT for segment P_s, that depends on the choice of pivot $p_{s,l}$ and on the assumed pivot match q_k.

For any non-pivotal element $p_{s,i}$, we may find several candidates on account of the relaxed inter-point distance criterion, or only a null-candidate, possibly due to occlusion or distortion above the threshold ε or wrongly assumed pivot match. For searching efficiency, the column of non-null candidates for a given $p_{s,i}$ is ordered by increasing value of the distance ratio in Eq.4, thereby moving the best candidates towards the column head.

In this way, each point $q_k \in \overline{Q}_M$ obtains a CT, but at most one of the tables (since the match point of $p_{s,l}$ may be lost) includes a correct match of segment P_s. With high probability, the CT with the correct match contains more candidates than other CTs. Therefore, to economize the search procedure, we discard some CTs with low numbers of candidates, and arrange priority of the rest according to the number of the candidates included.

If a join point in the segment P_s has already been identified during its parent-segment identification, reasonably this join point is used as the only pivot. In this case, only one CT is generated. This implies a large reduction of the search space.

3.1.2 CTs-based iterative matching

The CTs contain a number of candidates for each $p_{s,i}$. They effectively restrict the search space and make unnecessary the assumption of small motion or pose similarity required between two point-sets in the Iterative Closest Point (ICP) algorithm [3, 18]. In order to detect a correct match of segment P_s, we choose a CT in priority order and take the most reasonable candidates, $q'_{s,i} \in Q'_s$ say, to be an assumed correspondence with the pivotal sequence. If the set size $|Q'_s| \geq 3$, calculate the affine transformation $[R'_s, T'_s]$ by the SVD-based motion estimate algorithm [2, 9] related to this assumed correspondence. If the obtained motion estimation satisfies Eq.2, then Q'_s is regarded as a partial match of P_s. Otherwise, Q'_s has spurious matches and must be updated from the CT iteratively to carry out motion estimation until a partial match has been found. Next, confirm whether the partial match can be improved to give a whole best match Q_s which satisfies Eq.3 as well. This is implemented by using the closest-neighbor method to reassign a correspondence between the transformed model points $p'_{s,i}$ under $[R'_s, T'_s]$, and $\overline{Q}_M$ in Eq.5. This iterative procedure is applied to prioritized CTs until a whole best match Q_s of P_s is found if it exists. The identified points will be taken out of $\overline{Q}_M$ to remove them from further consideration. If no best match can be found, the segment search has failed. This results in some ambiguities for child-segment identification or may hint a wrong parent-segment identification, considered during the global hierarchical search procedure (see 3.2). We

summarize the CTs-based iterative matching procedure in Fig. 1, where $|Q'_s|$ denotes the number of non-null points in Q'_s.

Input: prototype P_s and unidentified $\overline{Q}_M$.
Loop L: for l =1 to $(1-\beta)M_s$: assign a pivot $p_{s,l}$
  Step1: Candidate-tables generation (3.1.1)
    Create an non-decreasing distance pivotal sequence;
    Find candidates (Eq.4) and generate CTs;
    Discard tables with fewest candidates;
    Arrange priority of remainders (biggest first);
    Sort candidates columns of each CT (best first).
  Step2: CTs-based iterative point matching (3.1.2)
    Loop A: input a CT. Do A until the last CT:
      Initialize Q'_s;
      Loop B: Finding partial match Q'_s: Do B while $|Q'_s| \geq 3$
        Update Q'_s from CT;
        Calculate motion by SVD-based algorithm;
        if $\bar{e}_s < \varepsilon$ (Eq.2)
          C: Finding a whole best match Q_s:
          Find closest neighbors (Eq.5);
          Update match Q'_s;
          Calculate motion;
          if $\bar{e}_s < \varepsilon$ and $|Q'_s| > \beta M_s$ (Eq.2,Eq.3)
            $Q_s \Leftarrow Q'_s$ as the best match of P_s;
            Remove the identified points from $\overline{Q}_M$;
            Return: segment match success.
      End Loop B
    End Loop A
End loop L: select another pivot $p_{s,l}$ in P_s.
Return: segment match failure.

Figure 1: CTs-based iterative matching

To update Q'_s from a CT in Fig.1, two methods are used. For a segment with more than a few points (e.g. our threshold is 6 for non-rigid human segments), we use the first candidate of each $p_{s,i}$ in the CT as initial match of the pivotal sequence to calculate an approximate transformation $[R'_s, T'_s]$. If this correspondence includes spurious pairs, the average matching error would not satisfy Eq.2. We update the initial match by replacing the worst match with its next candidate in the CT, based on the cue that matching errors of spurious pairs are higher than those of the correct pairs. This "coarse" update approach can quickly investigate the prioritized sequences of CTs to locate a CT satisfying a local match. This method has proved efficient for identifying the first segment, which has a large number of CTs to be investigated.

For a segment of only a few points or of poor rigidity, the above cue becomes unreliable, because the matching errors distribute more evenly so as to hide outliers. In this case we recourse to a nearest-furthest-to-pivot sequential search, which forms the second method. Each iteration uses only three points for the assumed partial match. The idea for choosing the pivot's nearest and furthest neighbors is that such points in a segment have

less ambiguity to distinguish them from others with intermediate distances [17].

To achieve a whole segmental match, the algorithm must find the remaining matches. Searching the wider set $\overline{Q}_M$ rather than the CT elements allows recovery even when a correct match has been dropped during the iteration or was not included in the CT because selection by 2ε in Eq.4 was strict, on grounds of limiting the CT search space. We apply the obtained motion estimate $[R'_s, T'_s]$ to all the model points in P_s: $p'_{s,i} = R'_s p_{s,i} + T'_s$, and find the closest-neighbor match $(p'_{s,i}, q_{s,i})$, $q_{s,i} \in \overline{Q}_M$, defined by Eq.5.

$$(p'_{s,i}, q_{s,i}) = \min_{q_i \in \overline{Q}_M} d(p'_{s,i}, q_i) < 2\bar{l}_s \varepsilon \tag{5}$$

If no such closest neighbor is found, we say the match point of $p_{s,i}$ is lost, and set $r_{s,i} = 0$.

3.2 Global hierarchical search strategy

Global articulated matching is achieved by integrating local segment-based matching with a hierarchical search strategy. In the segment-based articulated model, we assume that one of its segments contains more points and has more segments linked to it than most other segments. We treat such a segment as root. The global hierarchical search strategy begins at the root. After the root has been located, searching proceeds depth first to children along hierarchical chains. In this process, a joint may have been located in parent-segment identification, therefore it can be used as a known pivot in the child-segment. This linkage considerably increases the reliability and efficiency of the child-segment identification. In the case of a missing joint, an identified parent-segment of at least three points allows the motion transformation to recover a corresponding virtual joint, ensuring that child-segment identification can still proceed reliably.

When a segment has only few points, such as a point-pair model segment or a pose segment with enough missing points, local identification may be non-unique. In order to solve these kinds of uncertainties, we confirm such a segment in the hierarchical chain depending on whether its child-segment, even its grandchild-segment, can be found.

Failure to identify a child-segment may imply a wrong parent-segment identification. In this case, the algorithm attempts a backward error correction to the parent-segment. When a search chain in the hierarchy is broken by a failed segment identification, global searching will tend to identify other segments on other chains first and leave any remaining child-segments on broken chains the last to be solved.

4 Experimental Results

We have implemented the proposed SAPM algorithm in Matlab and applied it to both synthetic pose data and a real-world registration of human movement in 3D Moving Light Displays (MLDs). In our experiments, all model data and motion data are acquired from a marker-based optical MoCap Vicon 512 system. It includes 7 high-resolution calibrated cameras. The system can reconstruct the three-dimensional coordinates of an infrared retro-reflective marker if the marker is located in at least two camera views. The measurement accuracy of this system is at a level of a few millimetres in a control volume spanning meteors in linear extent.

4.1 Synthetic data

A number of model patterns have been investigated in our experiments. An example of a human model is shown in Fig.2, left. This model has 13 segments and 44 points in total; each segment contains up to 6 points, with the root segment (torso) having the most number. Segment-based matching started at the torso and spread to all the child-segments, such as head, pelvis, left/right arms and legs, on hierarchical chains.

In the first series of experiments, we study the ability of the proposed algorithm to identify arbitrary noise-free poses. To obtain such pose data, we apply a common random translation to all points in the model, then apply to each segment a random 3D-rotation around its joint. We attempt to identify the resulting synthetic configurations with the SAPM algorithm. Tests with about 2000 such synthetic poses show that correct identification occurs in about 95% of the cases. We show a typical case in Fig.2, in which each pair of identified feature points in the same segment is shown by stick links. We found the proposed SAPM algorithm works well without any similarity assumption on position and pose between the model and its observed data.

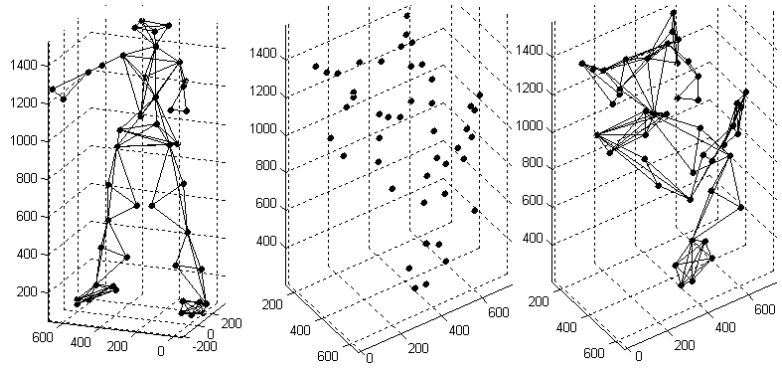

Figure 2: Identification with synthetic pose data: model (left), unidentified pose data (middle), identified pose (right).

The second series of experiments were carried out with segmental distortion applied to the configurations. For each point in a segment P_s, we added zero-mean Gaussian noise $N(0, \psi\bar{l}_s)/\sqrt{6}$, with a standard deviation scaled by distortion level ψ and the average segmental length $\bar{l}_s$, to its Cartesian coordinates x, y and z respectively. The tests included 5 configuration levels, in which the rotation angles applied to each segment ranged respectively from 0 to 4 radians of a uniform random distribution. Level 0 rotation denotes the same configuration as the model. The average identification rate versus increasing distortion and configuration level is given in Fig.3. The smoothed surface is based on the average of 10 random configurations on a given configuration level, with 100 distortion levels ψ in the range 0% to 10% of $\bar{l}_s$.

4.2 Real data

In this section, we report some results for identification of human poses from their freeform movements in 3D-MLDs. Human motion is a typical articulated motion with deformable segments. In these experiments, markers as external features were attached at key sites,

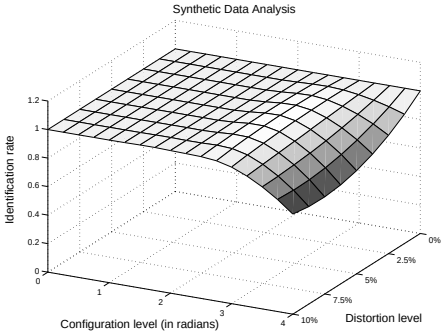

Figure 3: Identification rate versus distortion and configuration.

tightly clothed to limit distortion. A number of subjects, movements and feature point distributions were investigated. Illustrative results from a representative full-body marker protocol for a model with 15 segments and 49 feature points are shown in Fig.4. The model (Fig.4, left) is identified manually, using one frame of clear pose data. The observed point set can be taken from any frame of the captured motion data of the subject. The data may be subject to unexpected missing points, extra noise points and segment distortion among human body. From the example results in Fig.4, we observe the proposed algorithm is capable of identifying an articulated pose and dealing with the noisy data in some degree. Even when some key points, such as join points, have been lost, the algorithm can still carry on the child-segment identification successfully, referring to recovered virtual join points. For example, the lower limbs (second view in Fig.4) and head (third view in Fig.4) are successfully identified even with missing joints at the torso.

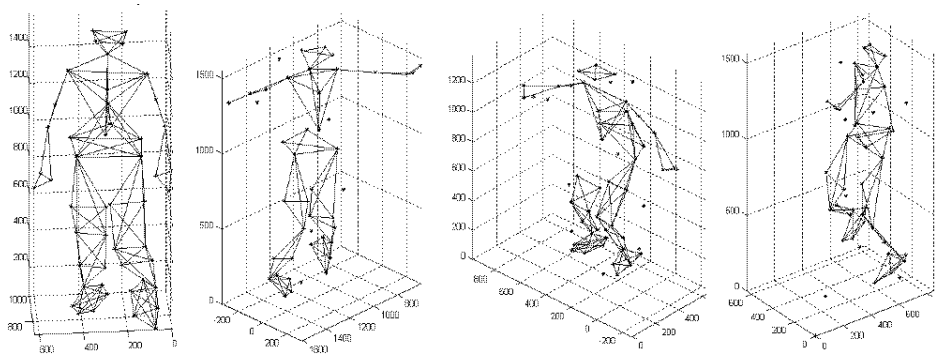

Figure 4: Identification of human pose in 3D-MLDs with missing and noisy data: model(left), identified poses of jumping(second), bending (third), running (fourth).

It is important to choose a set of appropriate parameter values for β and ε relevant to the type of articulated subjects and data quality. Generally, for subjects with rigid segments (robot manipulators), distortion tolerance ε in Eq.2 can be made small to raise the ability of rejecting outliers, and the matching size requirement β in Eq.3 can be low to allow for more missing data. For subjects with deformable segments (humans), we lower

the matching quality criterion to tolerate the raised distortion ε, but at the cost of bigger CTs and increased burden of search (Eq.4). That may also result in some false matches. In this case, we raise the matching size parameter β for compensation which, however, decreases the ability to handle missing data. In Table 1, we list parameter values used for experiments of human pose identification described above.

parameters	for actual human pose in 3D-MLDs
ε (in Eq.2)	10% $\sim$ 15%
β (in Eq.3)	80% $\sim$ 90%

Table 1: Parameters used for human pose identification of freeform movements.

The execution time depends not only on the number of points to be identified, but also on the level of distortion, the pose complexity and quality of the data. When we applied our algorithm on human motion data and executed Matlab code on a 866MHz Compaq with 256MB of RAM, the time for identifying a single frame of pose data of full-body human models with 30 to 50 points was around 3$\sim$5 seconds.

5 Conclusions

We have presented a new algorithm for articulated 3D sparse feature point matching to address the problem of an automatic initialization and accurate pose estimation in motion tracking in a best-fit sense. This aspect of initialization has received only limited attention. Most published works deal with incremental pose estimation and do not accommodate bootstrapping. We do not make the common simplifying assumptions, such as pose similarity or small motion, nor do we assume any kind of prior knowledge to infer an initial or partial correspondence between the two point-sets. It is the effectiveness of initialization that ultimately determines the robustness of the motion tracking. The algorithm remains a candidate for on-line initialization in point-feature motion tracking.

Acknowledgments:
All model data and motion data used in this paper were obtained by a marker-based optical MoCap system - Vicon 512, manufactured by Vicon Motion Systems Ltd., installed at the Department of Computer Science, UWA.

References

[1] J. K. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Articulated and elastic non-rigid motion: A review. In *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Object*, pages 2–14, Austin, TX, 1994.

[2] K. Arun, T. Huang, and S. Blostein. Least square fitting of two 3-D point sets. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(5):698–700, 1987.

[3] P. J. Besl and N. D. McKay. A method of registration of 3-D shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):239–255, 1992.

[4] L. Boxer. Even faster point set pattern matching in 3-D. In *Proc. SPIE Vision Geometry*, pages 168–178, 1999.

[5] C. Cédras and M. Shah. A survey of motion analysis from moving light displays. In *Proc. IEEE CVPR'94*, pages 214–221, Washington, June 1994.

[6] Z. Chen and H. Lee. Knowledge-guided visual perception of 3-D human gait from a single image sequence. *IEEE Trans. Sys., Man, Cybern.*, 22(2):336–342, 1992.

[7] L. Herda, P. Fua, R. Plänkers, R. Boulic, and D. Thalmann. Skeleton-based motion capture for robust reconstruction of human motion. In *Proc. Computer Animation*, Philadelphia, USA, May 2000.

[8] D. Lavine, B. Lambird, and L. Kanal. Recognition of spatial point patterns. *Pattern Recognition*, 16(3):289–295, 1983.

[9] A. Lorusso, D. W. Eggert, and R. B. Fisher. A comparison of four algorithms for estimating 3-D rigid transformations. In *Proc. British Machine Vision Conf.*, pages 237–246, Birmingham, UK, 1995.

[10] D. M. Mount, N. S. Netanyahu, and J. L. Moigne. Efficient algorithms for robust feature matching. *Pattern Recognition Special Issue on Image Registration*, 32:17–28, May 1998.

[11] S. Ranade and A. Posenfeld. Point pattern matching by relaxation. *Pattern Recognition*, 12:269–275, 1980.

[12] A. Rangarajan, H. Chui, E. Mjolsness, S. Pappu, L. Davachi, P. S. Goldman-Rakic, and J. S. Duncan. A robust point matching algorithm for autoradiograph alignment. *Medical Image Analysis*, 4(1):379–398, 1997.

[13] J. Richards. The measurement of human motion: A comparison of commercially available systems. *Human Movement Science*, 18(5):589–602, 1999.

[14] M. Ringer and J. Lasenby. Modelling and tracking articulated motion from multiple camera views. In *Proc. British Machine Vision Conf.*, pages 172–182, Bristol, UK, Sep. 2000.

[15] J. C. Sabel. *Calibration and 3D reconstruction for multicamera marker based motion measurement*. Delft University Press, Delft, Netherlands, 1999.

[16] Y. Song, L. Goncalves, E. D. Bernardo, and P. Perona. Monocular perception of biological motion - detection and labeling. In *Proc. IEEE ICCV*, pages 805–812, 1999.

[17] P. B. V. Wamelen, Z. Li, and S. S. Iyengar. A fast algorithm for the point pattern matching problem. *IEEE Trans. PAMI. last revised Dec.*, 1999.

[18] Z. Zhang. Iterative point matching for registration of free-form curves. *Computer Vision*, 3(2):119–152, 1994.