Efficient Marker Matching Using Pair-wise Constraints in Physical Therapy

Gregory Johnson, Nianhua Xie, Jill Slaboda, Y. Justin Shi,
Emily Keshner, and Haibin Ling

Temple University, Philadelphia, USA

Abstract. In this paper, we report a robust, efficient, and automatic method for matching infrared tracked markers for human motion analysis in computer-aided physical therapy applications. The challenges of this task stem from non-rigid marker motion, occlusion, and timing requirements. To overcome these difficulties, we use pair-wise distance constraints for marker identification. To meet the timing requirements, we first reduce the candidate marker labels by proximity constraints before enforcing the pair-wise constraints. Experiments with 38 real motion sequences, our method has shown superior accuracy and significant speedup over a semi-automatic proprietary method and the Iterative Closest Point (ICP) approach.

1 Introduction

Infrared tracked marker analysis is widely used for human motion analysis in computer-aided physical therapy [8] and related applications. In his paper, Klaus Dorfmuller-Uhaas uses a Kalman Filter to perform optical motion tracking [4]. Alexander Hornung discusses a method that automatically estimates all parameters on the fly [6]. Kazuutaka Kurihaha proposed an optical motion capture system with pan-tilt camera tracking, which expanded capturing range [9]. Victor B. Zordan used a physical model to map optical motion capture data to corresponding skeletal motion [15]. L. Herda performed studies for capturing skeletal motion as well [5]. Greg Welch et. al. introduced the Hi-Ball Tracking System [13]. This enabled Virtual Reality applications by generating over 200,000 head-pose estimates per second with very little noise and latency. Hirokazu Kato studied marker analysis for an application in augumented reality conferencing [7].

The key challenges are fast marker detection and fast processing. In this paper we first focus on marker identification problem, which is the first step of further marker sequence analysis. One way to model the marker identification problem is through marker tracking. A large amount of research effort has been conducted on this topic [14] in computer vision. The motion of all markers as a set is non-rigid since it articulates human motion. While rigid transformation approximates well for very small human movement, such as swaying, it is not suitable for the motion patterns in our task. Therefore, the non-rigidity brings difficulties to many marker tracking approaches such as the Kalman filter [2, 4]. Another solution is to track each marker independently and use essentially the local proximity to decide marker correspondences. This works fine when marker motions are small and reliable, but this condition is often violated in our study. More relevant technologies can be found in [14], such as [3, 20, 16].

Our goal is to provide a reliable and efficient solution to the marker sequence tracking/identification problem. Given the labeling of an initial frame, we render the problem as a point set corresponding problem and apply pair-wise distance constraints for makers from rigid parts of human body. We further improve the speed by restricting search of three nearest neighbors when forming candidate label sets. The proposed approach was tested on 38 sequences in comparison with the current semi-automatic proprietary system and the Iterative Closest Point (ICP) approach [1, 22]. The results show that our method not only significantly improves previous used semi-automatic system, but also runs much faster. Our method requires no manual labor except the labeling of the initial framework, which largely reduces the tedious human work. We also show the by showing our method provides more accurate results than an application of the ICP method.

2 Problem Formulation

We define a marker motion sequence as $S = \{I_t\}_{t=1}^{n}$, which contains n frames of marker positions. In a reference frame I_r, there are N identified marker positions and corresponding labels $l_1 = \{1, ..., N\}$ that are either manually or automatically labeled. In the rest of frames, markers are all un-identified. Let the t-th frame be $I_t = \{p_{t,i}\}_{i=1}^{N_t}$, $t \neq 1$, where N_t is the number of markers in I_t, $p_{t,i}$ is the position of the i-th marker. The identification task is to find a mapping $\pi : \{1, \dots, N_t\} \rightarrow \{0, 1, \dots, N\}$, such that $p_{t,i}$ and $p_{1,\pi(i)}$ are from the same marker if $\pi(i) > 0$. Our task has the following challenges:

- **Missing markers**. There are often some markers missing due to occlusion or system errors.
- **Ghost markers**. Spurious markers sometimes appear, which do not correspond to any marker labels. These markers are called *ghost markers* due to their unpredicted spatial and time appearance. The ghost markers are caused by signal detection errors and the dropping of markers during human motion.
- **Efficiency**. Currently, using a propriety semi-automatic procedure with a commercial system, it takes a post doctoral researcher several hours to annotate a two-minute captured marker sequence.

We now use a toy example (shown in Fig.1) to illustrate the matching process and our solution. We let $P = \{p_1, p_2, \dots, p_N\}$ be the positions of N identified makers at time $t - 1$ and let $Q = \{q_1, q_2, \dots, q_N\}$ be the positions of N un-identified makers at time t. Examples with $N = 7$ are shown in Fig. 1 (a) and (b).

Many previous systems used in physical therapy study build the mapping $\pi(.)$ through nearest neighbor matching. In other words, the solution intends to reduce the following cost:

$$C(\pi) = \sum_{i=1}^{N_t} |p_{t,i} - p_{r,\pi(i)}|^2. \tag{1}$$

Such a solution, while efficient, is problematic with the presence of missing and ghost markers. We developed a second order cost method to improve the matching.

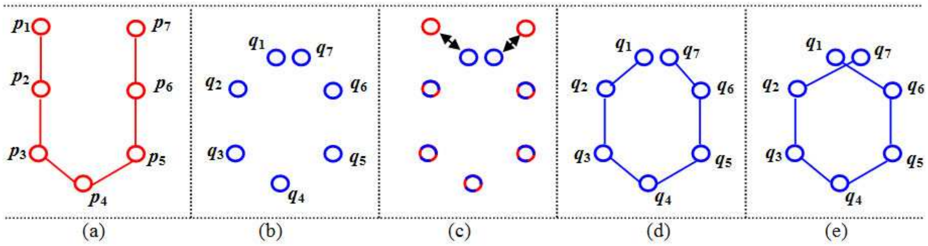

Fig. 1. Pair-wise distance constraints for marker identification. (a) An identified marker set P at frame t-1. The links between points show the skeleton of two arms viewed from the top of the skeleton, such that p_1 to p_3 are markers on the left arm and where p_5 to p_7 are markers on the right arm. (b) The marker set Q at frame t, which is to be matched to P. (c) Matching results without consider pair-wise distance. (d) Resulted skeleton from (c). (e) Resulted skeleton from the proposed matching.

Intuitively, we can include in (1) the deformation constraints from all pairs of markers. Such a direct solution is very expensive and practically un-necessary. It is natural to restrict the constraints in selected pairs of markers that reflect human motion structures, e.g., the links between a hand and an elbow, but not the link between the head and an ankle. Denote $E = \{(i, j)\}$ as the set of such links, we now extend (1) as following:

$$C(\pi) = \sum_{i=1}^{N_t} |p_{t,i} - p_{r,\pi(i)}|^2 + = \lambda \sum_{(i,j)\in E} c(p_{t,i}, p_{t,j}; p_{r,\pi(i)}, p_{r,\pi(j)}), \qquad (2)$$

where λ is the regularization weight, $c(.)$ is the cost of matching segment $(p_{t,i}, p_{t,j})$ to $(p_{r,\pi(i)}, p_{r,\pi(j)})$. Intuitively, the first term on the right hand side models "proximity constraints" and the second term models "geometric constraints", which is defined by the deformation between pairs of markers. A natural selection of $c(.)$ is the absolute difference between Euclidean distances $|p_{t,i}p_{t,j}|$ and $|p_{r,\pi(i)}p_{r,\pi(j)}|$, that is:

$$c(p_{t,i}, p_{t,j}; p_{r,\pi(i)}, p_{r,\pi(j)}) = ||p_{t,i}p_{t,j}| - |p_{r,\pi(i)}p_{r,\pi(j)}||. \qquad (3)$$

For efficiency, we pre-compute the pairwise distance matrix D from the reference frame as $D_{i,j} = |p_{r,i} - p_{r,j}|$. An example reference frame with annotation is shown in Fig. 2.

Directly minimizing the cost function is very expensive. Instead, we turn to find a heuristic solution by taking into account first the proximity constraints and then the geometric constraints.

3 Algorithm

Our task is to find the matching $\pi(.)$ from current frame I_t to the reference frame I_r. Motivated by the above discussion, we propose a two-stage solution for finding the matching $\pi(.)$, followed a step to update the reference frame p_r.

The first stage addresses the proximity constraint, i.e., $\sum_{i=1,\ldots,N_t} |p_{t,i} - p_{r,\pi(i)}|^2$. For this purpose, we build *candidate labels* $\mathcal{C}_i \subset \{0, 1, \ldots, N\}$ for each point $p_{t,i} \in I_t$.

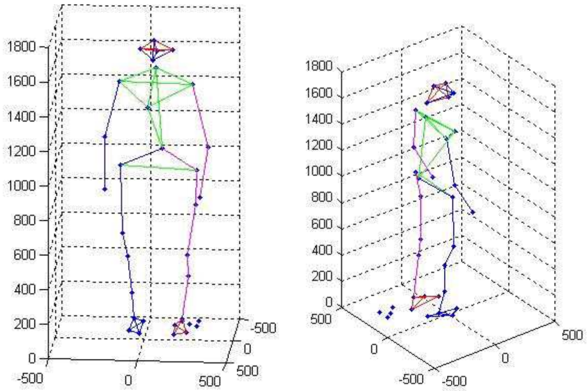

Fig. 2. An example frame with manual annotation. Each line represents a pair-wise constraint

Specifically, for each $p_{t,i}$, we first choosing the three markers from I_r that are closest to it. Then further check these three candidates based on their absolute distances from $p_{t,i}$ and how they related to each other. After this step, the complexity of our matching task is largely reduced to picking the best candidate from $\mathcal{C}_i$, which contains at most three candidates. The candidate set $\mathcal{C}_i$ may contain 0–3 candidate labels for point $p_{t,i}$. Our problem then contains two tasks: reducing the ambiguity for $|\mathcal{C}_i| > 1$ and solving matching conflicts. Both tasks are dealt in the second stage.

The second stage addresses the geometric constraint, i.e., $\sum_{(i,j)\in E} c(p_{t,i}, p_{t,j}; p_{r,\pi(i)}, p_{r,\pi(j)})$. We use the pairwise constraints for this purpose. In particular, for each candidate label $i' \in \mathcal{C}_i$, the distances from $p_{r,i'}$ to its linked neighbors should be similar to the distances from $p_{t,i}$ and its potential neighbors. This heuristic solution is effective in the physical therapy application, since high frame rate is used and the variation are either very small (for true correspondence) or are fairly large (for ghost points).

After these two stages, we also need to update the reference frame. This is based on fusing the previous reference frame with the new matching result. One challenge here is caused by missing markers, especially continuous missing markers. Fortunately, the pairwise constraints provide again reliable way to recover them. The basic idea is that, for a missing marker, the positions of its neighbors can provide a strong constraint for its position. The details of the whole algorithm is summarized in Algorithm 1. In the algorithm, the thresholds τ_k, $k = 1, 2, \ldots, N$ are used for determine ghost points. We use $\tau_k = 0.25$ for all labels except for $\tau_6 = 0.4$, $\tau_9, \tau_{21} = 0.35$. The three exceptions are for elbows and T-1 bone's positions, which are usually unstable due to large human motion and system errors.

4 Experimental Results

4.1 Database

We use a human motion database including 38 sequences, 27 of which come from non-patients and the rest come from patients. The sequences consist of the subjects taking a

Algorithm 1 Automatic Marker Matching

```
**Input**: The reference frame $I_r$ and the frame $I_t$.
$\mathcal{U} \leftarrow \varnothing$.
**for** $i = 1..N_t$ **do**
    For $p_{t,i}$, find its three nearest markers in $I_r$, denote their labels as $l_1$, $l_2$, $l_3$.
    Calculate their distances $d_1, d_2, d_3$ to $p_{t,i}$. Without loss of generality, $d_1 \leq d_2 \leq d_3$.
    **if** $d_1 < d_2/4$ **then**
        $\mathcal{C}_i \leftarrow \{l_1\}; \mathcal{U} \leftarrow \mathcal{U} \bigcup \{l_1\}$
    **else**
        **if** $d_1 \geq d_2/4$ and $d_1 < d_3/4$ **then**
            $\mathcal{C}_i \leftarrow \{l_1, l_2\}$
        **else**
            $\mathcal{C}_i \leftarrow \{l_1, l_2, l_3\}$
        **end if**
    **end if**
**end for**
Remove duplicates in $\mathcal{U}$
**for** $i = 1..N_t$ **do**
    $\pi(i) \leftarrow 0$,     /*default, ghost point*/
    **if** $|\mathcal{C}_i| == 1$ **then**
        $\pi(i) \leftarrow \mathcal{C}_i(1)$
    **else**
        **for** $k \in \mathcal{C}_i$ **do**
            **for** $p_{t,j}$ whose label $j \in \mathcal{U} \bigcap E_k$, where $E_k = \{m : (m,k) \in E\}$ **do**
                $e_{k,j} = ||p - p_{t,j}| - D_{k,j}|/D_{k,j}$
            **end for**
            $e_k = \max_j \{e_{kj}\}$
        **end for**
        $e = \min_k \{e_k\}$
        $l = arg\min_k \{e_k\}$
        **if** $e < \tau_l$ **then**
            $\pi(i) \leftarrow l$
        **end if**
    **end if**
**end for**
Update the reference frame $I_r$ accordingly.
```

few steps forward or backward, as well as sitting down or standing up. Each sequence contains approximately 7800 frames. All sequences are 120 frames per second. Each marker is labeled manually by an expert using semi-automatic software (Motion Analysis from Santa Rosa, CA) for comparison. As shown in Fig.2, the ground truth number of markers is 34 for all sequences, but because of missing or ghost markers, the number of markers in a frame varies from 28 to 36. Our algorithm only uses one frame as the manually labeled frame, which may contain ghost markers.

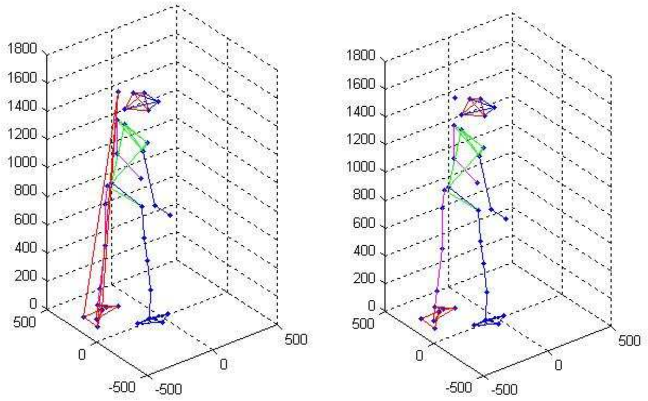

Fig. 3. A frame with different identifications. Left: results from semi-automatic labeling. Right: results of our method

The sequences from non-patients usually have rare missing markers but may contain ghost markers, while the patient sequences often contain lots of missing and ghost markers. Therefore, the patient sequences are more difficult and interesting.

4.2 Comparison with Manual Marker Labeling

We compared the results of our algorithm with those from manual labeling using a semi-automatic commercial software. We compare the results frame-by-frame. We are mainly interested in frames where the two methods generate different labels. For the 27 non-patient videos, all the identification results are the same. This means that our method works as well as Motion Analysis. Using our algorithm eliminates the human intervention and is much faster (a few minutes as opposed to a few hours of manual labeling).

For the 11 patient videos, which are more pertinent within the context of physical therapy applications, our method generates different label configurations from the semi-automatic system in 2333 frames. By analyzing these frames one-by-one carefully, we find that whenever our method disagrees with the previous solutions, it is either due to the incorrect labeling of the previous system or due to the ambiguity in maker positions. Specifically, among the 2333 frames, our predictions are correct in 2259 frames, and the remaining 74 frames have ambiguous labels. One example is shown in Fig. 3. For example, in some frames, two markers may be very close in proximity to each other. Also, a marker may be largely disturbed in one axis for several frames.

In summary, the experimental results show that our method outperformed the commercial system with manual marker labeling. In addition, our method runs about 70 seconds per sequence with current Matlab implementation. This is extremely fast compared with previous solutions that usually take an expert five to six hours.

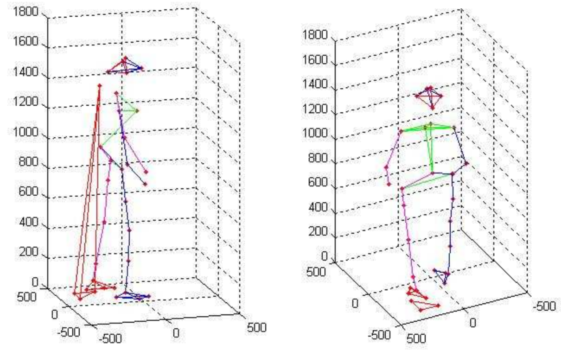

Fig. 4. A frame with different identifications. The marker near T-1 bone should be labeled 6(T-1 bone) as right figure showed, however, it's incorrectly labeled 34 by human

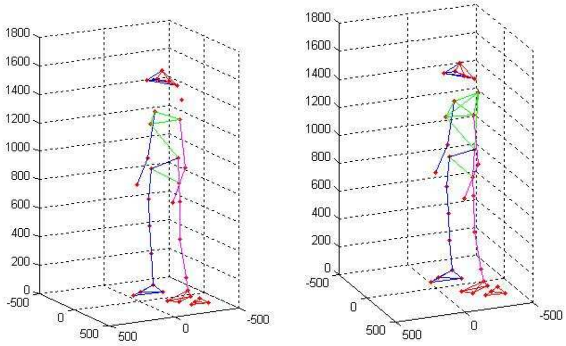

Fig. 5. A frame with different identifications. The marker near T-1 bone should be labeled 6 as shown in the right figure, however, it is incorrectly labeled as a ghost marker

4.3 Comparison Without Pairwise Constraints

We also compared our method with a method without pairwise constraints. The proposed method certainly outperforms nearest principal method when the frame is far away from human-labeled frame. This means that the inner distance constraints play an important role in matching.

4.4 Comparison with Iterative Closest Point Algorithm

The Iterative Closest Point Algorithm [1, 22] is used to minimize the difference between two point clouds. The algorithm takes in two groups of points and outputs the transformation parameters between them. The first step is to associate the points between the groups by nearest neighbor criteria. Next, the transformation parameters are estimated using a mean square cost function. The points are then transformed using the estimated

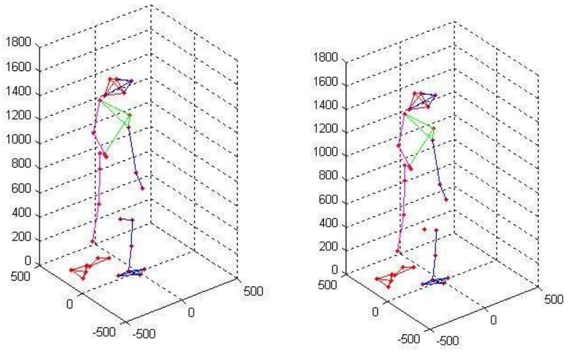

Fig. 6. A frame with different identifications. The marker near right leg should be a ghost marker as right figure showed, however, it is incorrectly labeled 13—right thigh

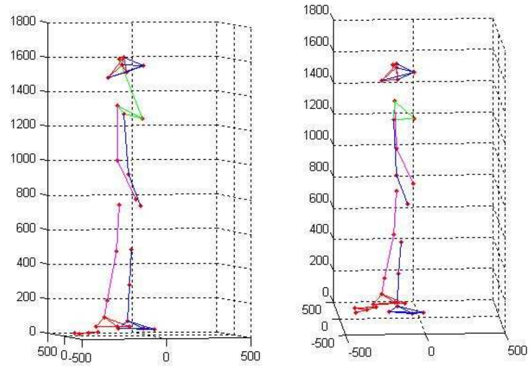

Fig. 7. A frame with different identifications. The marker near head should be a ghost marker as right figure showed, however, it is incorrectly labeled as 11—L4/L5

parameters, and the process is repeated a number of times based on a predetermined stopping criteria.

We can use this method as a comparison tool to our algorithm. Using ICP, we can go frame by frame and determine the point in the previous frame that corresponds to each point in the current frame. Using an initially labelled frame, we can find the corresponding points in the next frame, and then label these points with the same labels as in the previous frame. We repeat these process with the entire sequence, using the most recently labeled frame as the new ground truth each time. After this process, we will have an array of frame label predictions in a similar format as the results produced in our algorithm.

Using fourteen different sequences, we used both the ICP algorithm and our algorithm to predict labels for all frames, given one initially labeled frame for each sequence. We then compared the labelling results for each frame from the two algorithms. If the results of a frame differed, a diagram of both algorithms' labels for that frame was

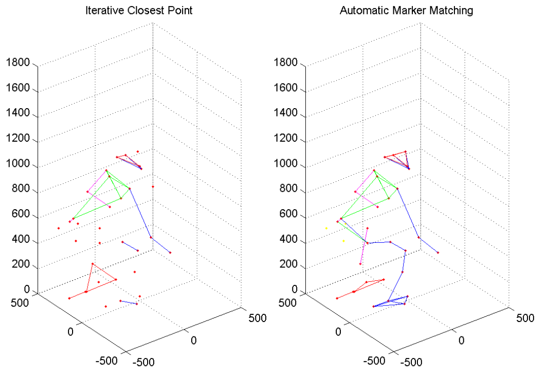

Fig. 8. An example comparison between the Iterative Closest Point algorithm and the Automatic Marker Matching algorithm. In this frame, the Automatic Marker Matching algorithm is more effective

saved as an image file. After going through all sequences, we went through each of the images created to see which algorithm produced a better set of labels for that frame. We have a correctly labeled initial frame for each sequence, so we know the basic structure to look for when analyzing these files. Figure 8 is an example of these image files.

Of the 19,620 frames found, the Automatic Marker Matching algorithm performed better in 17,842 of them (90.94%), while the ICP algorithm performed better in 689 frames (3.51%). In 1,089 frames (5.55%), the visual difference was too small to declare one algorithm more effective than the other.

5 Conclusion

In this paper we reported a fast and robust algorithm for automatic infrared tracked marker identification for physical therapy applications. Our method uses both proximity and pairwise constraints. Experiments showed that our method not only generates better accuracy than the current commercial system, but also runs much faster. We expect to apply the reported method to production runs and exploit other potential motion analysis applications.

Acknowledgment This work is supported in part by NSF Grant IIS-1049032, NIA-NIH grant AG26470. Gregory Johnson is supported by a fellowship supported by NSF Grant DGE-0841377.

References

1. Besl, P. J. and McKay, N. D. “A Method for Registration of 3-D Shapes”, *PAMI*, 14:239-256, 1992.

2. T. J. Broida and R. Chellappa, ”Estimation of Object Motion Parameters from a Sequence of Noisy Images”, *PAMI*, 8:90-99, 1986.
3. M. Isard, A. Blake, ”Condensation - conditional density propagation for visual tracking”, *IJCV*, 29:5-28, 1998.
4. K. Dorfmuller-Ulhaas, ”Robust optical user motion tracking using a kalman filter”, *ACM VRST*, 2003.
5. L. Herda, P. Fua, R. Plankers, R. Boulic, and D. Thalmann, ”Skeleton-based motion capture for robust reconstruction of human motion”, *Proc. Computer Animation*, 2000.
6. A. Hornung, S. Sar-Dessai, and L. Kobbelt, ”Self-calibrating optical motion tracking for articulated bodies”, *IEEE Virtual Reality*, pp. 75-82, 2005.
7. H. Kato, and M. Billinghurst, ”Marker tracking and HMD calibration for a video-based augmented reality conferencing system”, *Int’l Wshp on Augmented Reality*, pp. 85-94, 1999.
8. E. A. Keshner and R.V. Kenyon. ”Using immersive technology for postural research and rehabilitation”, *Assist Technol. Summer*, 16(1) : 54-62, 2004
9. K. Kurihara, S. Hoshino, K. Yamane, and Y. Nakamura, ”Optical motion capture system with pan-tilt camera tracking and real time data processing”, *ICRA*, vol. 2, 2002.
10. R. van Liere, and A. van Rhijn, ”Search space reduction in optical tracking”, *Proceedings of the workshop on Virtual environments*, pp. 207-214, 2003.
11. M. Ringer, and J. Lasenby, ”A procedure for automatically estimating model parameters in optical motion capture”, *Image and Vision Computing*, vol. 22, pp. 843-850, 2004.
12. D. Tolani, A. Goswami, and N. I. Badler, ”Real-time inverse kinematics techniques for anthropomorphic limbs”, *Graphical models*, vol. 62, pp. 353-388, 1999.
13. G. Welch, G. Bishop, L. Vicci, S. Brumback, and K. Keller, ”The HiBall tracker: High-performance wide-area tracking for virtual and augmented environments”, *ACM VRST*, 1999.
14. A. Yilmaz, O. Javed, and M. Shah, ”Object Tracking: A Survey,” *ACM Computing Surveys*, Vol. 38, No. 4, 2006.
15. V. B. Zordan, and N. C. Van Der Horst, ”Mapping optical motion capture data to skeletal motion using a physical model”, *ACM symp. on Computer Animation*, 245-250, 2003.
16. V. Salari and I. K. Sethi. ”Feature point correspondence in the presence of occlusion.” *PAMI*, 12(1):87-91, 1990
17. I. Sethi and R. Jain. ”Finding trajectories of feature points in a monocular image sequence”, *PAMI*, 9(1):56-73, 1987
18. K. Rangarajan and M. Shah. ”Establishing motion correspondence”, *Conference Vision Graphies Image Process 54*, 1, 56-73, 1991.
19. S. Intille, J. Davis, and A. Bobick. ”Real-time closed-world tracking.” *CVPR*, 697-703, 1997.
20. C. Veenman, M. Reinders, and E. Backer, ”Resolving motion correspondence for densely moving points.” *PAMI*, 23(1):54-72, 2001.
21. K. Shafique and M. Shah, ”A non-iterative greedy algorithm for multi-frame point correspondence”, *ICCV*, 110-115, 2003.
22. Z. Zhang, “Iterative point matching for registration of free-form curves and surfaces.” *IJCV*, 13:119-152, 1994.