Hand motion expertise analysis using dynamic hierarchical activity modeling and isomap

Kanav Kahol*^, Mithra Vankipuram*
Human machine symbiosis laboratory,
**Department of Biomedical Informatics, School of Computing and Informatics, Arizona State University, Phoenix, Arizona*
^Sim. Education and Training Center, Banner Good Samaritan Medical Center, Phoenix, AZ
{kanav; mithra.vankipuram}@asu.edu

Abstract

Several domains such as sports, surgery, dance etc. are characterized by a significant influence of expertise of the performer on the motion pattern and style. The retrieval of expertise level of the performer through automated motion analysis is the subject of this paper. We employ a novel neuropsychologically inspired algorithm that employs a dynamic hierarchical layered structure to represent the human anatomy, and low-level parameters to characterize motion in the layers of this hierarchy which correspond to different segments of the human body. This characterization is representative of the expertise of the performer of the motion. These motion profiles are then compared using dynamic time warping to generate a similarity matrix. We employ isomap to reduce dimensionality and the cluster the data into different expertise classes. This algorithm was tested on a library of surgical movements that contained 3D hand motion data of common surgical laparoscopic procedures. Linearly separable clusters were obtained between novice, intermediate and expert performances. Test sequences were projected into manifold spaces. A recognition percentage of 98.56% was obtained for classifying the test sequences into correct expertise clusters.

1. Introduction

An intriguing and a challenging area for automated human motion analysis lies in the detection of quality of motion or in other words the expertise of the performer of motion. Domains such as sports, dance, motor rehabilitation and even surgery can benefit from development of automated algorithms that can retrieve the expertise of a performer from low level motion features [1, 2]. These automated systems can serve as a basis for educational purposes providing valuable information to the user about their motion profiles, how it differs from that of an expert and how to improve the same. Several such systems are available for sports analysis, rehabilitation [1-3] etc. These systems tend to attribute expertise to control of a few motion parameters in the human body. For example golf analysis tends to focus exclusively on wrist movement analysis and all other features of motion, like the back swing and follow through are ignored [4]. Such types of systems are very beneficial for focused movement analysis. However there are several types of movement patterns that involve movement of several segments of the human body interacting in a complex dynamic manner and such reductionist approaches may not provide an adequate basis for modeling the complexities of the human motion. Research has shown that humans understand movement in the human body as a whole rather than as fragmented interactions of joints and segments. Hence, algorithms that can model human motion in a holistic manner will not only facilitate better expertise analysis but also provide better feedback for the performers. With efficient use of pattern recognition techniques it is possible to conduct this type of analysis.

The expertise recognition algorithm proposed in this paper uses (1) a dynamic hierarchical layered structure to model the human body and (2) activity measures in human body segments to determine the expertise class of a subject. We employ the isomap technique to reduce the dimensionality of temporal activity vector profiles that capture the dynamic interactions between body segments to reveal a manifold space separating experts, intermediate performers and novices.

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

2. Related Work

Automated motion analysis to determine expertise has been the subject of several studies, especially in sports analysis. In the rehabilitation arena, the problem is reformulated as that of distinguishing between normal motion and pathological motion patterns. There are several commercial and academic systems available for such analysis. McCrory and Berkovic [2] developed a video analysis based system to model acute motor and convulsive manifestations in sports related concussions. Rostkowska et al. [4] developed a system called AVIImage which allowed analysis of movement, by means of video registration with RGB 24 bits depth. Churchill et al. [1] developed a video based system that tracked visual markers to reveal gait patterns. All these systems use mechanisms like tracking for basic analysis and then analyze kinematic features through either simple statistical means [2] or pattern recognition [1]. Two types of approaches are usually used to characterize human motion sequences through pattern recognition. (1) State-space approaches, which use points, line and 2D/3D blobs, and (2) template matching approaches, which use overlaid meshes to characterize particular movement patterns. Different models are generally defined for different classes of expertise. The key problem however remains in identifying representative low level features. This is often a time consuming step and requires experts to visually annotate features that are salient. As a result, many of these systems are analysis systems and not direct classification systems.

Isomaps have been recently applied for motion analysis especially gesture recognition and activity recognition. Li et al. applied Isomaps to CMU's gait image database [5] to classify the type of walk (as slow walk, fast walk, slow incline walk or slow walk holding a ball). Results included 1d and 2d isomap embedding. Blackburn and Rebeiro employed isomap and dynamic time warping for activity recognition [6]. They employed isomaps to reduce dimensionality of individual synchronized frames and then added the time dimensions. Dynamic time warping was employed to match embedded manifolds to test sequences. This approach was applied for activity recognition which is a different problem compared to expertise recognition. Expertise recognition requires analysis of the same type of motion but with different parameters. This makes the problem inherently more complex than activity recognition which has more inter-class variation than expertise wherein subtle differences in motion cause differences in expertise.

3. Hierarchical layered template of the human hand

At a mechanical level, human hand can be modeled as a complex system of hierarchically connected rigid units as in Figure 1. The kinematic model given here has the palm divided into upper palm (including the ulner border and hypothenar eminence) and the thenar eminence which controls the thumb [7]. Each segment can move independently, and hence can exhibit an independent degree of activity. These segments can be represented as a hierarchy as shown in figure 1. The lowest layer of this hierarchy is called layer 1, the next higher layer is called layer 2, etc. Parent segments in each layer can have child segments in the layer below theirs. A parent segment inherits the aggregate characteristics of all of its child segments. For example, the momentum of a thenar eminence will be a vector sum of the momentum of thumb proximal phalanx and the thumb distal phalanx. Child segments inherit the motion of the parent segment. Thus the distal phalanx of the thumb inherits the motion of the thenar eminence, although it might also have a motion of its own.

In motion sequences there might be some time periods during which two adjacent segments have very similar motion vectors. When this happens these two segments are perceived as a single segment. Also if the relative orientation of two adjacent segments does not change over a period of time, they are perceived as a single segment. However, most adjacent segments will move with some perceptible degree of independence, and will thus be perceived as separate segments. Consider for example the gesture of exploring the texture of a curved surface through the index finger. Initially the finger establishes touch with the surface during which the index finger moves as a whole but distinct segments also show independent movements. However when the lateral movement of the index finger is started then the entire finger moves as one unit (see inset in Figure 1). It should be noted that this process takes place dynamically. Thus, the number of single motion sequence, and the corresponding body hierarchy diagram will also be dynamic.

4. Hierarchical activity modeling for expertise analysis

In this section we define a generic algorithm to model hand movements as a function of activity of individual segments of the hierarchy as shown in Figure 1.

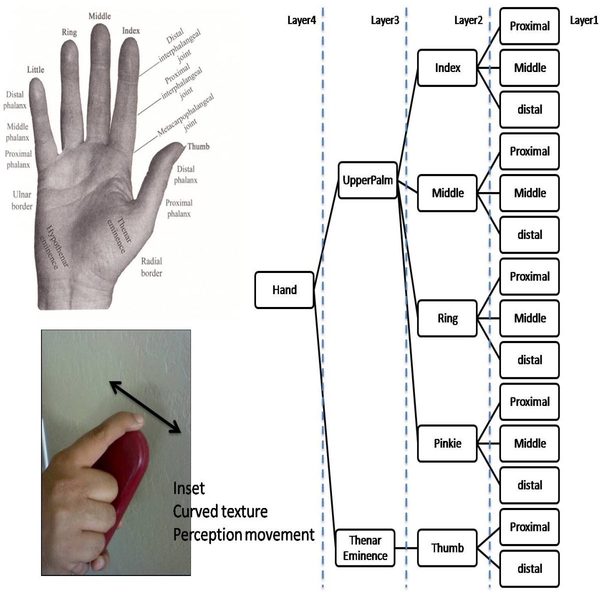

Figure 1. Dynamic hierarchical model of hand

The SegmentActivity of each segment can be computed from three motion parameters associated with each segment – namely momentum, Kinetic energy and Force.

F(SegmentKE, SegmentForce, SegmentMomentum) (4.1)

Kinetic energy of a segment and momentum of a segment requires an estimation of the velocity of the segment and the segment force requires estimation of the acceleration. These estimates can be formed by tracking algorithm or 3D motion capture. All the terms also require calculation of the mass of each individual segment. This is estimated from ergonomic equations that provide proportional estimates of masses of the hand regions and segments. The idea behind using momentum, energy and force metrics is to employ a weighting mechanism for the propagation of motion parameter values from the lower layers to the upper layers of the hierarchy. For example activity of proximal phalange of the thumb should contribute more to the activity of thenar imminence as it is larger and thicker in size. If unweighted velocity and acceleration are employed, they do not contribute to expertise analysis. Higher Layer Segment Force, Segment Momentum, and Segment Kinetic Energy are computed by taking the vector sum (or in some cases the scalar sum) of all their respective child segments. Through this methodology a measure of activity in every segment of the human body can be defined. We employed Gaussian smoothing (low-pass filtering) to produce a smoother measure of activity that correlated better with motion expertise.

For each frame of the motion sequence the activity triple is calculated for each of the segments. In addition to the triples we also included information on the dynamic characteristic of the hand hierarchy in each frame. An ordered vector was included to depict pairs of the segments in the hierarchy that have coalesced to form a unified segment. Coalescing of segments is determined by the following process. An initial configuration of the hand is calculated as the mean and standard deviation of the angles between segments. This is generally calculated at the beginning of the data capture. During the sequence if the angle remains within one standard deviation from the mean then the segments are deemed to coalesce else they behave as disparate segments. This simple algorithm leads to determination of the binary vector depicting a 1 for pairs of segments that were behaving as a single unit and 0 otherwise. It may be recounted that our model allows for hierarchical interactions.

5. Isomap based expertise Analysis

Activity triples and the dynamic hierarchical vector form adequate basis for detection of expertise of motion. For the next step, we require to reduce the dimensionality of the motion sequence. We chose isomap based manifold learning as our technique as expertise patterns are inherently nonlinear but can be represented as embedded manifolds in a lower dimensional space. As a first step complex motion sequences are divided into simpler units employing a gesture segmentation algorithm. Kahol et al. presented a system for gesture segmentation based on activity profiles which was employed for this purpose [8]. The individually segmented gestures are compared with other gestures to yield a pairwise distance measure. In order to compare the gesture sequences we employed dynamic time warping (DTW). DTW allows for efficiently comparing two sequences that may differ in length. As our algorithm works with simple gestures rather than complete motion sequences the requirement of monotonicity was met.

Let $Y = \{y_i \in \mathrm{R}_d,\ i = 1, \ldots, N\}$ be the set of preprocessed gesture sequences and $X = \{x_i \in \mathrm{R}_m,\ i = 1, \ldots, N\}$ be the corresponding embedding points. The embedded points X are determined using the following three-step Isomap algorithm: **(1)** Create a weighted graph G of points in Y with weights $d_Y(i, j)$ representing the pairwise distance between neighbors. In our algorithm, a neighborhood is defined by the k-nearest neighbors; **(2)** Estimate the pairwise geodesic distances $d_G(i, j)$ between all manifold points by finding the shortest path distances in the graph G. These shortest path distances are denoted by $d_G(i, j)$. We chose geodesic distance as informal experiments showed best results for the geodesic distance measure. **(3)** Finally, we apply classical multidimensional scaling (MDS) on D_G to map the data onto an m-

dimensional Euclidean space X. It is worth pointing out that $d_X(i, j)$ and $d_Y(i, j)$ are Euclidean pairwise distances within manifold space while $d_G(i, j)$ represents the actual geodesic distances. The coordinate vectors $\mathbf{x}i$ in X are chosen by minimizing the stress function given by

where is in an operator that converts distances to inner products as described. Mapping new sequences into an embedded manifold requires an invertible mapping between the original space and the embedded manifold space. We employed the algorithm proposed by [9] that is based on radial basis functions. For expertise recognition of new sequences in the manifold space, we find the Euclidean distance between test sequence and individual cluster centers. This distance is used to find the proximity of the sequence to each cluster. The sequence is then assigned to the closest cluster.

6. Experiments and Results

We applied the proposed algorithm on a library of 600 laparoscopic sequences as measured through a Cyberglove® with 22 sensors. These data capture sequences represent surgeons performing common laparoscopic movements on validated surgical simulators. Laparoscopic movements are complex in their nature. The hand movement data was captured at 120 frames per second. Initial motion sequences were segmented into gestures. Each gesture was labeled as expert, intermediate or novice. The activity triples and the dynamic hierarchical vector was calculated for each frame and smoothed using temporal Gaussian smoothing. DTW was employed to compare motion sequences against each other to yield pairwise distances. The pairwise distances werethen converted to geodesic distance as outlined above. Classical MDS was applied to reduce the dimensionality. We employed 400 sequences for obtaining the initial clusters and then 200 labeled sequences were used as test sequences. Figure 2 shows the cluster space in 2 dimensions as obtained. It can be seen that the obtained clusters were linearly separable. 97% of the test sequences were correctly classified when the clusters were defined by the k-means algorithm with k=3. The test sequences were projected into the manifold space using the algorithm from [9]. Test sequences were classified with an accuracy of 98.56%.

7. Conclusions

Dynamic hierarchical activity modeling and isomap technique can suitably model hand motion for expertise analysis. In future, this work will be extended to include whole body movements. The hierarchy can be modified as in [8] to achieve this. The formulation is not specific to a sensing modality or the classes of hand motion allowing for generality of the proposed framework. We also propose to test this algorithm to other classes of hand motion such as tennis etc. Finally the technique will also be tested with video data which is available in the dataset.

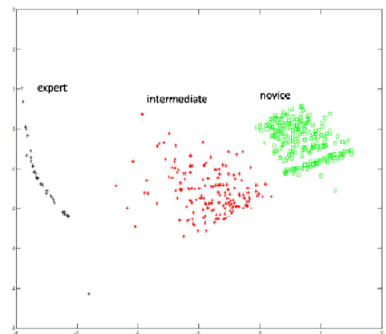

Figure 2. Clusters of expertise in the scaled spaces

8. References

[1] A. Churchill, P. W. Halligan, and D. T. Wade, "RIVCAM: a simple video-based kinematic analysis for clinical disorders of gait " *Comp. Methods and Programs in Biomedicine*, vol. 69, pp. 197-209, 2001.

[2] P. R. McCrory and S. F. Berkovic, "Video analysis of acute motor and convulsive manifestations in sport-related concussion," *Neur.*, v. 54, pp 1488-1491, 2000.

[3] J. K. Aggarwal and Q. Cai, "Human Motion Analysis: A Review," *Comp. vision and image u.*, v. 73, pp. 428-440, 1999.

[4] E. L. ROSTKOWSKA, B. PRZEMYSŁAW, and D. B. LECHOSŁAW, "AVIImage – video motion analysis software for tests of biomechanical movement characteristics," *Acta Bioeng. & Biomech.*, v. 8, 2006.

[5] H. Li and X. Li, "Gait analysis using ISOMAP," presented at Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on, 2004.

[6] J. Blackburn and R. E., "Human Motion Recognition Using Isomap and Dynamic Time Warping," in *Human Motion- Understanding, Modeling, Capture and Animation: Lecture Notes in Computer Science*, vol. 4814/2007. Berlin: Springer, 2007, pp. 285-298.

[7] L. A. Jones and S. J. Lederman, *Human Hand Function*. New York: Oxford, 2006.

[8] K. Kahol, P. Tripathi, and S. Panchanathan, "Gesture Segmentation in Complex Motion Sequences," presented at IEEE International Conference on Image Processing, Barcelona, Spain, 2003.

[9] A. Elgammal and C.-S. Lee, " Inferring 3D body pose from silhouettes using activity manifold learning," presented at IEEE Int. Conf. on Comp. Vision and Pattern Recognition, Los Alamitos, CA, 2004.