Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

2D Subspaces for Sparse Control of High-DOF Robots

Odest Chadwicke Jenkins

Abstract— **We investigate the use of five dimension reduction and manifold learning techniques to estimate a 2D subspace of hand poses for the purpose of generating motion. Our aim is to uncover a 2D parameterization from optical motion capture data that allows for transformation sparse user input trajectories into desired hand movements. The use of shape descriptors for representing hand pose is additionally explored for dealing with occluded parts of the hand during data collection. We present early results from uncovering 2D parameterizations of power and precision grasps and their use to drive a physically simulated hand from 2D mouse input.**

I. INTRODUCTION

Human interfaces for controlling complex mechanical/robotic systems, such as humanoid robots and mechanical prosthetic arms, is an underdetermined problem. Specifically, the amount of information a human can reasonably specify is far less than the variables for controlling a robot system, or degrees-of-freedom (DOFs). Basic control tasks for humans, such as reaching and grasping, are often onerous for human operators of robot systems. Such robot control problems persist even for able bodied human operators given state-of-the-art sensing and actuation platforms, such as the NASA Robonaut [1].

These problems become magnified for applications to biorobotics, particularly those applications related to prosthetic and assistive devices by users with lost physical functionality. In such applications, feasible sensing technologies, such as electroencephalogram (EEG) [2], electromyography (EMG) [3], and cortical neural implants [4], [5], provide a very limited channel for user input due to the sparsity and noise of the sensed signals. Efforts to decode these user signals into control signals have demonstrated success limited to 2-3 DOFs. An upper boundary on the bandwidth of neural interfaces could be liberally estimated as 10 bits/sec. With such limited bandwidth, applications for these decoding efforts have been focused on low-DOF systems, such as 2D cursor control [6] and mobile robots [2].

Robotic systems geared for general functionality or human anthropomorphism will have significantly more than 2-3 DOFs. For instance, a prosthetic arm and hand could have around 30 DOF. While this mismatch in input and control dimensionality is problematic, it is clear that the space of valid human arm/hand poses does not fully span the space of DOFs due to biomechanical redundancy. Statistically, it is very likely that plausible configurations exist in a significantly lower dimensional subspace [7], [8], [9]. In general, uncovering intrinsic dimensionality of this subspace is critical for bridging the divide between decoded user input and producing robot control signals. A related question of interest is whether there are 2D subspaces of hand poses suitable for use with existing decoding methods. In other words, how can we remap the control of high dimensional robotic systems into low dimensional spaces to get the most out of current decoding methods.

In this paper, we investigate the use of various dimension reduction techniques to estimate a 2D subspace of hand poses for the purpose of generating hand motion. Our aim is to uncover a 2D parameterization that transforms sparse user input trajectories into desired hand movements. Our current work focuses on *data-driven* methods for analyzing power and precision grasps obtained through optical motion capture. We apply five dimension reduction techniques for embedding this data in a 2D coordinate space: Principal Components Analysis (PCA) [10], Hessian LLE [11], Isomap [12], Windowed Isomap, and Spatio-temporal Isomap [13]. The use of shape descriptors for representing hand pose is additionally explored for dealing with occluded parts of the hand during data collection. We present early results from uncovering 2D parameterizations of power and precision grasps and their use to drive a physically simulated hand from 2D mouse input.

II. APPROACH

Our objective for this work is to estimate a control mapping $f : Y \rightarrow X$ that maps 2-dimensional control space $y \in \Re^2$ into the space of hand poses $x \in \Re^d$. d is the number of DOFs expressing hand pose. The estimation of the mapping f is founded upon the assumption that the space of plausible hand poses is intrinsically parameterized by a low-dimensional manifold subspace. We assume each hand pose achieved by a human is an example generated within this manifold subspace. It is given that the true manifold subspace of hand poses is likely to have dimensionality greater than two. With an appropriate dimension reduction technique, however, we can preserve as much of the intrinsic variance as possible. As improvements in decoding occur, the dimensionality of the input signal will increase but we will still leverage the same control mapping.

We create a control mapping by taking as input a set of training hand poses $x_i \in \Re^d$, embedding this data into control space coordinates $y_i \in \Re^2$, and generalizing to new data. The configuration of points in control space $y_i = f^{-1}(x_i)$ is latent and represents the inverse of the control mapping. Dimension reduction estimates the latent coordinates y such that distances between datapairs preserve some criteria of similarity. Each dimension reduction method

O.C. Jenkins is with the Department of Computer Science, Brown University, Providence, RI, 02912-1910, USA `cjenkins@cs.brown.edu`

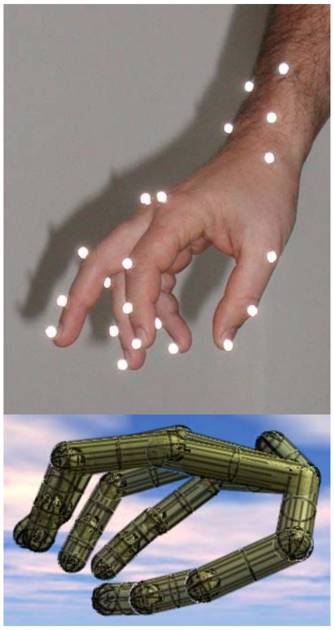

Fig. 1. (top) Snapshot of a hand instrumented with reflective markers for optical motion capture. (bottom) Our physically simulated hand attempting to achieve the captured hand configuration above.

has a different notion of pairwise similarity and, thus, a unique view of the intrinsic structure of the data. Once embedded, the input-control pairs (x_i, y_i) are generalized through interpolation to allow for new (out-of-sample) points to be mapped between input and control spaces.

A. Dimension Reduction Overview

Our method relies upon a suitable notion of pairwise similarity for performing dimension reduction. We explored the use of five dimension reduction methods for constructing control mappings. The first method, PCA, constructs a linear transform of rank 2 A about the mean of the data μ:

$$y = f^{-1}(x) = A(x - \mu) \tag{1}$$

PCA is equivalent to performing multidimensional scaling (MDS) [10] with pairwise Euclidean distance. MDS is a procedure that takes a full matrix D, where each element $D_{x,x'}$ specifies the distance between a datapair, and produces an embedding that attempts to preserve distances between all pairs. MDS is an optimization that minimizes the "stress" E of the pairwise distances:

$$E = \sqrt{\sum_{x} \sum_{x'} (\sqrt{(f(x) - f(x'))^2}) - D_{x,x'})^2} \tag{2}$$

The second method, Isomap, is MDS where shortest-path distances are contained in D as an approximation of geodesic distance:

$$D_{x,x'} = \min_p \sum_i D'(p_i, p_{i+1}) \tag{3}$$

where D' is a sparse graph of local distances between nearest neighbors and p is a sequence of points through D' indicating the shortest path. Isomap performs nonlinear dimension reduction by transforming the data such that geodesic distance along the underlying manifold becomes Euclidean distance in the embedding. A canonical example of this transformation is the "Swiss roll" example, where input data generated by 2D manifold is contorted into a roll in 3D. Given a sufficient density of samples, Isomap able to flatten this Swiss roll data into its original 2D structure, within an affine transformation. Isomap assumes input data are unordered (independent identical) samples from the underlying manifold, ignoring the temporal dependencies between successive data points.

Windowed Isomap accounts for these temporal dependencies in a straightforward manner by windowing the input data $\tilde{x}_i = x_i \dots x_{i+w}$ over a horizon of length w. This augmentation of the input data effectively computes local distances D' between trajectories from each input point rather than individual data points.

ST-Isomap further augments the local distances between the nearest neighbors of windowed Isomap. ST-Isomap reduces the distance between the local neighbors that are spatio-temporal correspondences, i.e., representative of the k best matching trajectories, to each data point. Distances between a point and its spatio-temporal correspondences are reduced by some factor and propagated into global correspondences by the shortest-path computation. The resulting embedding attempts to register these correspondences into proximity.

All of these methods assume the structure of the manifold subspace is convex. Hessian LLE [11] avoids this convexity limitation by preserving local shape. Local shape about every data point is defined by an estimate of the Hessian $H_{x,x'}$ tangent to every point. $H_{x,x'}$ specifies the similarity of x to neighboring points x'; non-neighboring points have no explicit relationship to x. This sparse similarity matrix $\tilde{H}_{x,x'}$ is embedded in a manner similar to MDS, where global pairwise similarity is implicit in the local neighborhood connectivity.

III. Results

The dimension reduction methods mentioned above were applied to human hand movement collected through optical motion capture. The performer's hand was instrumented with 25 reflective markers, approximately .5cm in width, as shown in Figure 1. These markers were placed at critical locations on the top of the hand: 4 markers for each digit, 2 for the base of the hand, and 3 on the forearm. A Vicon motion capture system localized each of these markers over time.

The performer executed several power and precision grips with different orientations of the hands. The resulting dataset consisted of approximately 3000 frames of 75 DOF data (25 markers in 3D). Markers that dropout (occluded such that their location cannot be estimated) were assigned a default location during intervals of occlusion. Embeddings produced by each of the dimension reduction methods are shown in

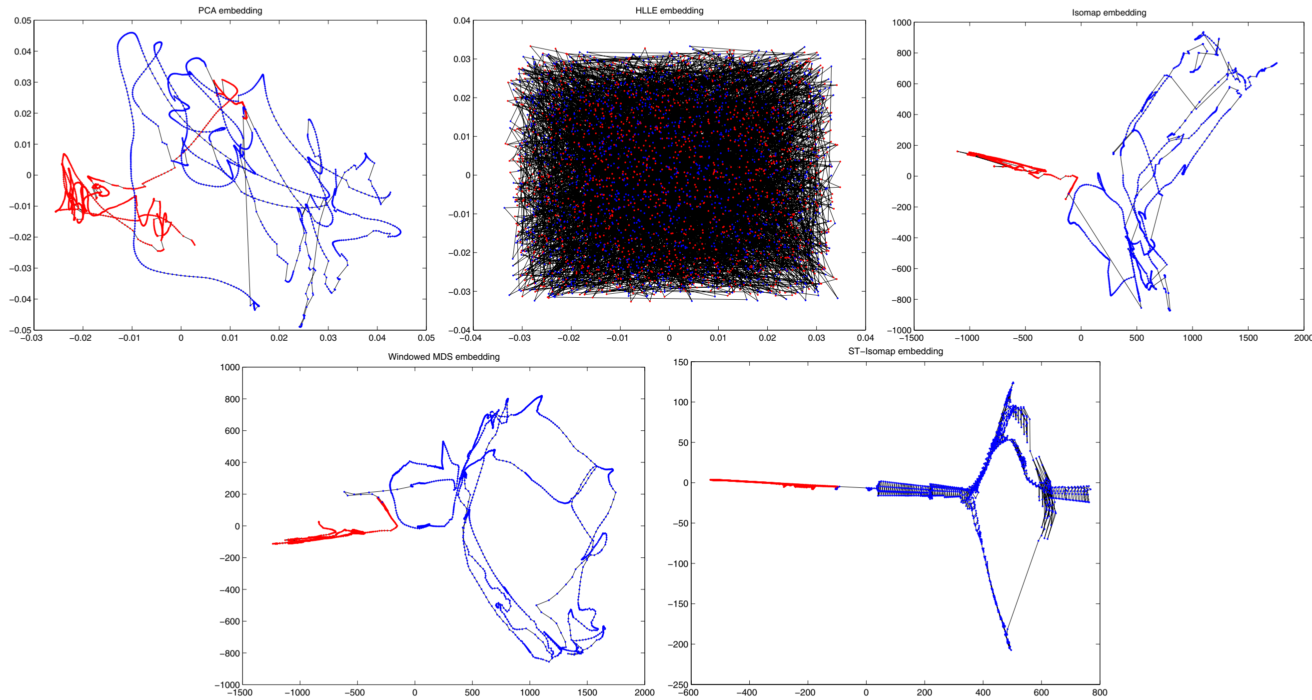

Fig. 2. Embedding of the hand marker positions using (in order) PCA , Hessian LLE, Isomap, windowed Isomap, and ST-Isomap. Power and precision grasp configurations are shown in blue and red, respectively. A black line is drawn between temporally adjacent configurations.

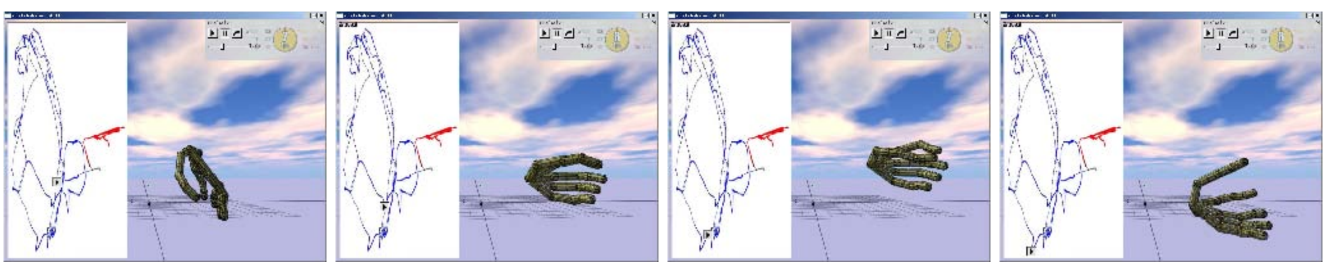

Fig. 3. A visualization controlling a power grasp for a physically simulated hand using 2D mouse control. A mouse trajectory in the windowed Isomap embedding (shown by the cursor position) is mapped into hand configurations. This trajectory shows an initial idle configuration, preshaping, grasp closure, and rotation.

Figure 2, with neighborhoods $k = 5$ for HLLE, $k = 50$ (Isomap), $k = 10$ (windowed Isomap), $k = 3$ (ST-Isomap). Windowing occurred over a 10 frame horizon. ST-Isomap reduced distances by a factor of 10 between points that are the $k = 3$ best local correspondences.

The quality of each embedding was assessed based on the amount of inconsistency, embedding locations that merge disparate hand configurations, and continuity over time. Inconsistency is an undesired artifact that is related to residual variance (or stress) not accounted for in the embedding. The affect of residual variance differs across methods and is only of consequence with respect to inconsistency. We were unable to find a reasonable embedding for HLLE, suggesting the space of marker configurations does have continuous local shape. The PCA embedding best preserves the consistency among the precision grasps. However, this embedding has many temporal discontinuities due to marker dropout and inconsistencies for power grasps. The Isomap embedding suffers from the same discontinuities with fewer inconsistencies of the power grasp and greater inconsistencies of the precision grasp. The windowed Isomap embedding removes the discontinuities of the Isomap embedding by effectively performing a square wave low-pass filter on the input before shortest-path computation. The ST-Isomap embedding was prone to overregistration of the data, which we attribute to a lack of adaptivity in the number of correspondences across local neighborhoods.

Although we found the windowed Isomap embedding to be the most promising, none of these methods appeared to be directly applicable to yield reasonable control spaces. We used the windowed Isomap embedding to control a physically simulated hand. The kinematics of the hand were created manually using the marker positions from a single frame in the data set. Cylindrical geometries were used to represent the hand. The Open Dynamics Engine was used to integrate physics over time. The desired configuration of

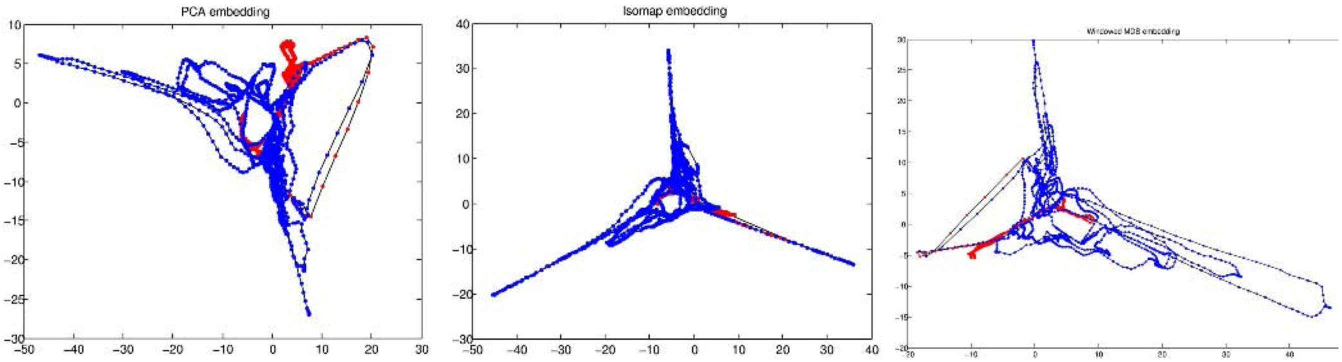

Fig. 4. Embedding of the centralized moments for hand marker positions using (left-to-right) PCA, Isomap, and windowed Isomap. Power and precision grasp configurations are shown in blue and red, respectively. A black line is drawn between temporally adjacent configurations.

the hand is set by the nearest neighbor in the 2D embedding of the current mouse position. This procedure allowed for reasonable control of some power grasps, such as in Figure 3. Control over precision grasps often cause the system to be unstable due to violent changes in the desireds from a small mouse movement.

Marker dropout from occluded parts of the hand was likely to be a large factor in the embedding inconsistency and discontinuity. To account for marker dropout, statistical shape descriptors were used to recast the input space with features less sensitive to specific marker positions. Centralized moments [10] up through the 16th order were computed for the non-occluded markers at every frame of the motion. The resulting data set contained 3000 data points with 4096 dimensions and was used as input into PCA, Isomap (k=20), and windowed Isomap (k=5). As shown in Figure 4, all three of these methods produced embeddings with three articulations from a central concentrated mass of poses, have better consistency between power and precision grasps, and exhibit no significant discontinuities. From our preliminary control tasks, the PCA embedding provides the best spread across the 2D space with comparable levels of consistency to windowed Isomap.

IV. Conclusion

We have explored the use of dimension reduction for estimating 2D subspaces for controlling high DOF robotic systems. Our aim in this work has been to provide a means for controlling high dimensional systems with sparse 2D input with an emphasis towards leveraging current decoding capabilities. We have taken a data-driven approach to this problem by using dimension reduction to embed motion performed by humans into 2D. Five dimension reduction methods were applied to motion capture data of human performed power and precision grasps. Centralized moments were utilized towards embedding with greater consistency and utilization of the embedding space. An analytical measurement of inconsistency and continuity with comparison to manually crafted control mappings and robustness to input noise will be performed in future work. Additionally, greater exploration of shape descriptors suitable for embedding hand poses offers an avenue for improved control space embeddings.

V. Acknowledgments

The author thanks Morgan McGuire, Michael Black, Greg Shaknarovich, and Mijail Serruya for simulation assistance and helpful discussions.

References

[1] W. Bluethmann, R. Ambrose, M. Diftler, E. Huber, A. Fagg, M. Rosenstein, R. Platt, R. Grupen, C. Breazeal, A. Brooks, A. Lockerd, R. A. Peters, O. C. Jenkins, M. J. Matarić, and M. Bugajska, "Building an autonomous humanoid tool user," in *IEEE International Conference on Humanoid Robots (Humanoids 2004)*, Nov 2004.

[2] J. del R. Millan, F. Renkens, J. Mourino, and W. Gerstner, "Brain-actuated interaction," *Artif. Intell.*, vol. 159, no. 1-2, pp. 241–259, 2004.

[3] M. Zecca, S. Micera, M. C. Carrozza, and P. Dario, "Control of multifunctional prosthetic hands by processing the electromyographic signal," *Critical Reviews in Biomedical Engineering*, vol. 30, no. 4-6, pp. 459–485, 2002.

[4] J. Donoghue, A. Nurmikko, G. Friehs, and M. Black, "Development of neural motor prostheses for humans," *Advances in Clinical Neurophysiology (Supplements to Clinical Neurophysiology)*, vol. 57, 2004.

[5] D. Taylor, S. H. Tillery, and A. Schwartz, "Information conveyed through brain control: Cursor versus robot," *IEEE Trans. Neural Systems Rehab Eng.*, vol. 11, no. 2, pp. 195–199, 2003.

[6] M. Serruya, A. Caplan, M. Saleh, D. Morris, and J. Donoghue, "The braingate pilot trial: Building and testing a novel direct neural output for patients with severe motor impairment," in *Soc. for Neurosci. Abstr.*, 2004.

[7] J. Lin, Y. Wu, and T. Huang, "Modeling the constraints of human hand motion," in *IEEE Workshop on Human Motion*, 2000.

[8] E. Todorov and Z. Ghahramani, "Analysis of the synergies underlying complex hand manipulation," in *Intl. Conference of the IEEE Engineering in Medicine and Biology Society*, 2004.

[9] C. R. Mason, J. E. Gomez, and T. J. Ebner, "Hand synergies during reach-to-grasp," *The Journal of Neurophysiology*, vol. 86, no. 6, pp. 2896–2910, Dec 2001.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.

[11] D. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proc. National Academy of Sciences*, vol. 100, no. 10, pp. 5591–5596, 2003.

[12] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[13] O. C. Jenkins and M. J. Matarić, "A spatio-temporal extension to isomap nonlinear dimension reduction," in *The International Conference on Machine Learning (ICML 2004)*, Banff, Alberta, Canada, Jul 2004, pp. 441–448.