Modeling the Product Manifold of Posture and Motion

Ankur Datta   Yaser Sheikh   Takeo Kanade

Robotics Institute

Carnegie Mellon University

{ankurd,yaser,tk}@cs.cmu.edu

Abstract

Long-term human motion is composed of an ensemble of different activities with varying complexity. This makes it challenging to develop models to accurately estimate human motion. In this paper, we exploit the dependencies that exist between posture and motion for long-term human motion estimation. We propose to model the nonlinear motion manifold as a collection of local linear models, noting that given a particular posture, the variation in motion for that posture can be well-approximated by a linear model. A collection of local linear models is easy to fit and also has the expressiveness to encode several activities in any arbitrary order. Furthermore, to account for the varying complexity of different activities, each local linear model can have a different dimensionality. A collection of local linear models, thus, avoids the limitation of global models that require a uniform dimensionality for the latent motion manifold. This model allows us to linearly regularize motion estimation algorithms over the nonlinear human motion manifold. Our results demonstrate that a collection of local linear models provides an effective representation for the motion manifold when compared to other global models such as the bilinear model [18] and the Principal Component Analysis [14].

1. Introduction

Human motion is influenced by the internal intentions of the actor and the external constraints of the actor's environment. Consider a driver inside a vehicle. The series of activities performed by the driver depends on what the driver wants to do (e.g. turn on the air conditioner, turn up the volume, or adjust the seat height), and what driving requires (e.g. shift gears, adjust the rear view mirror, and turn the steering wheel). The challenge in estimating motion in such settings is that both the intentions of the driver and the nature of environmental constraints usually remain hidden. As a result, human motion manifests itself as a series of activities with varying complexity whose order is arbitrary. In previous literature, activities have been modeled *globally*
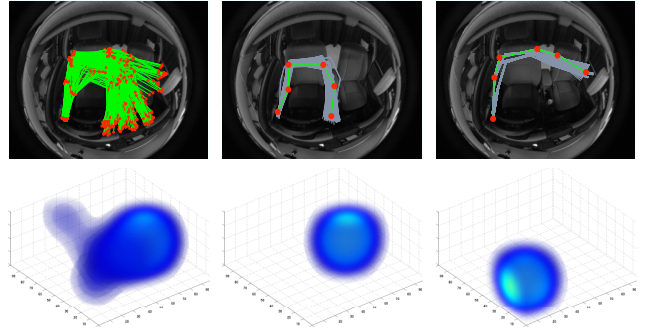

Figure 1. First column: top row – 700 posture exemplars out of a data set of 18,000 posture exemplars and bottom row – 3D iso-contour for 18,000 posture exemplars. Second and third column: tow row – Posture variability around two key postures and bottom row – their associated 3D iso-contours. Developing global models for the nonlinear iso-contour (first column) generated by the entire data set is a difficult task. However, by exploiting the local dependencies between posture and motion, we can compactly model the nonlinear manifold as a collection of Local Linear Models (LLMs) (second and third column).

by low-dimensional manifolds. However, to handle arbitrary orderings of activities, extending global models would require a combinatorial increase in the training data.

In this paper, we propose an approach that exploits local dependencies between posture and motion for human motion estimation. Given a *particular* posture, the variation in motion for that posture is compactly approximated using a local linear model. Taken together, a collection of local linear models can encode long-term human motion that may consist of human activities in any arbitrary order. Figure 1 illustrates the dependency between posture and motion by comparing a global model, fit to motion across a sequence of 18,000 frames, to local models, fit to motion given particular postures. It is evident from the figure that motion is highly constrained given knowledge of posture. A key benefit of our approach is that each local linear model can have a different dimensionality, thereby, well-approximating the irregular motion manifold. The irregularity of the motion manifold arises from the fact that dif-

ferent activities have different complexity. Allowing varying dimensionality across the manifold avoids the limitation of global models that require commitment to a uniform dimensionality.

2. Related Work

Humans are capable of performing a wide-variety of complex motion due to the highly articulated nature of the human body. Human motion data, therefore, resides in a high-dimensional configuration space. Previous approaches have attempted to model the high-dimensional motion manifold using either the linear models or the nonlinear models [6, 7, 16, 22, 20, 1, 10].

Linear motion models employ Principal Component Analysis (PCA) [14] to develop models for human motion in computer vision and graphics [13, 17, 1]. While it is straightforward to fit the linear models, the complexity of human motion does not lend itself particularly amenable to modeling using the linear models [20].

Nonlinear motion models such as the Mixture of Probabilistic PCA [20], Locally Linear Embeddings [11] or the Gaussian Process Dynamical Models (GPDM) [22], enable one to capture the nonlinearities in human motion at the expense of local-minima prone fitting. Nonlinear motion models can be divided into two categories, those that commit to a single manifold dimensionality to model the motion manifold [11, 10, 22] or those that model the motion manifold using a mixture of several local linear or nonlinear models such as Mixture of Probabilistic PCA or Geometric Topograhphic Mapping (GTM) [20, 19, 3]. Long-term human motion is an irregular manifold with a diverse range of activities with varying complexity (see Figure 6). Committing to a single dimensionality for the latent manifold, therefore, neither captures the manifold structure effectively nor extracts the computational savings [11, 22].

Nonlinear approaches that model the motion manifold as a mixture of local linear or nonlinear models suffer from high subspace angles between the learnt local models [20, 19, 3, 2, 8]. This problem was highlighted by the work of Roweis *et al.* [15, 21]. These high subspace angles between the local models arise because the traditional approaches employ a disjoint partition and fit approach which essentially makes the linear models independent causing high subspace angles between subspaces. Roweis *et al.* and Brant [15, 21, 4] outlined model selection algorithms that imposed a nonlinear regularization penalty to encourage neighboring subspace models to align.

In this paper, we propose an alternative approach to encourage neighboring subspace models to align. We propose to learn a collection of local models by increasing the data overlap between the neighboring models. Increased data overlap between the models automatically forces them to align with each other without the need for costly minima-
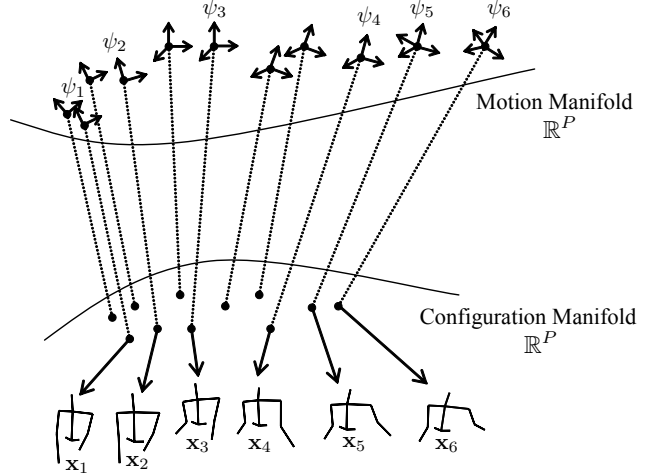

Figure 2. Product Manifold of posture and motion. Postures on the configuration manifold have associated Local Linear Models (LLMs) that encode the motion variation in the neighborhood of that particular posture.

prone regularization penalty. Our approach, however, does require comparatively more number of local models. The computational cost of doing so, though, is minimal as the local linear models are learnt using PCA over small neighborhoods on the motion manifold.

3. The Product Manifold of Posture & Motion

The posture $\mathbf{x} \in \mathbb{R}^P$ of a human is defined as the instantaneous configuration of a human body, in 2D image space or 3D world space. All human configurations $\mathbf{x}$ lie on a manifold $\mathbf{X}$. The instantaneous change in human posture is represented by $\dot{\mathbf{x}} = f(\mathbf{x}), f(\cdot) = \frac{d\mathbf{x}}{dt}, \dot{\mathbf{x}}_i \in \mathbb{R}^P$. The purpose of this paper is to estimate human motion. This is challenging since humans are highly articulate objects and move in complex, and unpredictable ways. If we could characterize the motion manifold $\dot{\mathbf{X}}$ then we could regularize the estimates of motion to lie on the manifold $\dot{\mathbf{X}}$. A P-dimensional *chart* on $\dot{\mathbf{X}}$ is a pair $(\mathbf{V}, \psi)$ consisting of a subset $\mathbf{V}$ of $\dot{\mathbf{X}}$ and a bijection ψ from $\mathbf{V}$ onto an open set in $\mathbb{R}^P$. $\mathbf{V}$ is the domain of the chart. A point $\dot{\mathbf{x}} \in \mathbf{V}$ can thus be assigned coordinates $(a^1, \ldots, a^P)$ by means of the projection mappings $\mathbf{v}^i : \mathbb{R}^P \to \mathbb{R}; a^i = \mathbf{u}^i \circ \psi, i = 1, \ldots, P$. Henceforth, we will denote a chart only by its bijection ψ when there is no ambiguity.

The human motion manifold is highly nonlinear and lies in a high-dimensional configuration space. It is therefore, a difficult task to find a single chart covering the entire manifold $\dot{\mathbf{X}}$. Linear or Non-Linear Dimensionality Reduction techniques (NLDR) attempt to model the high-dimensional manifold by computing a globally valid low-dimensional mapping into a latent coordinate system, usually the real-plane $\mathbb{R}^{P'}$, where $P' \ll P$. However, the operating as-
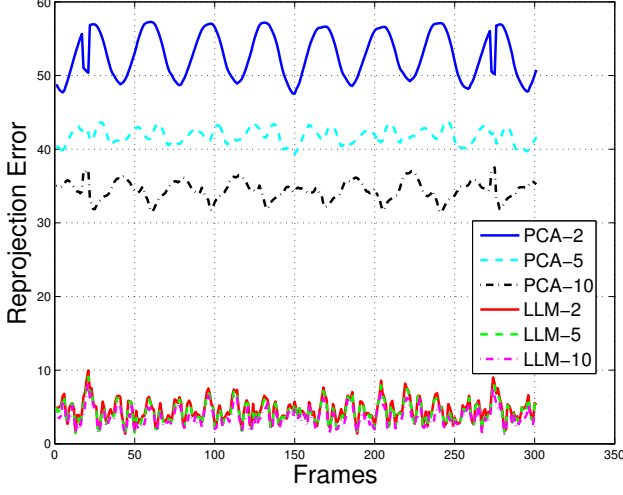
Figure 3. Comparing the representational power of PCA and a collection of Local Linear Models (LLMs). This graph shows the reconstruction error for held-out motion vectors from a motion capture data set of 70,000 frames. It is evident that 2, 3, and 10-dimensional PCA embedings of the motion manifold do far-worse than even a 2-dimensional embedding using a collection of LLMs.

sumption of such approaches is that a high-dimensional nonlinear manifold is globally *homoemorphic* to the real-plane $\mathbb{R}^{P'}$ is generally not true for manifolds like $\dot{\mathbf{X}}$. Figure 3 shows a comparison of representational power between the Principal Component Analysis (PCA) and the proposed collection of Local Linear Models (LLMs). We collected 70,000 frames of motion capture data of several people performing a variety of activities such as artistic walking, running, hopping etc. and learnt a model for the entire motion manifold $\dot{\mathbf{X}} \in \mathbb{R}^{39}$ using PCA and LLMs. It is evident that a 10-dimensional embedding using PCA does far-worse than a 2-dimensional embedding from a collection of LLMs, thereby, illustrating that LLMs are a much more faithful model for the motion manifold compared to PCA.

Since a single chart cannot effectively cover the entire manifold $\dot{\mathbf{X}}$, a collection of charts whose domain, taken together, cover $\dot{\mathbf{X}}$ are introduced. An *atlas* $\mathcal{A}$ is a collection $\{(\mathbf{V}_i, \psi_i)\}_{i \in I}$ of charts on $\dot{\mathbf{X}}$, I being an index set and $\psi_i(\mathbf{V}_i)$ an open set in $\mathbb{R}^P$, such that

1. $\dot{\mathbf{X}} = \bigcup_{i \in I} \mathbf{V}_i$,

2. For each pair $i, j \in I$ for which $\mathbf{V}_i \bigcap \mathbf{V}_j \neq \oslash$, the mapping $\psi_j \circ \psi_i^{-1}$ is a C^k diffeomorphism of $\psi_i(\mathbf{V}_i \bigcap \mathbf{V}_j)$.

In this paper, we model the motion manifold $\dot{\mathbf{X}}$ as a collection of locally linear charts. Each locally linear chart ψ_i can be estimated using a linear basis computed from motion exemplars that lie in the domain $\mathbf{V}_i$ of a particular motion exemplar $\dot{\mathbf{x}}_\mathbf{i}$. Traditionally, approaches that use a collection
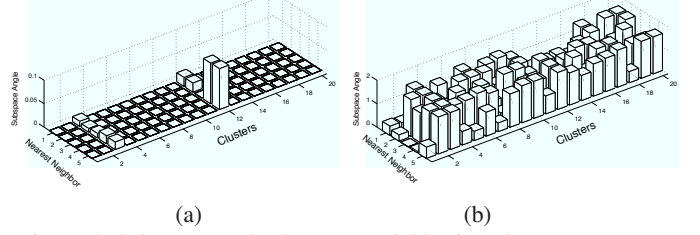

Figure 4. Subspace angles between neighboring charts. (a) Low subspace angles are obtained using the proposed approach that uses shared data to learn the charts. (b) High subspace angles are obtained using the Expectation-Maximization fitting of a Mixture of Probabilistic Principal Components Analysis [20].
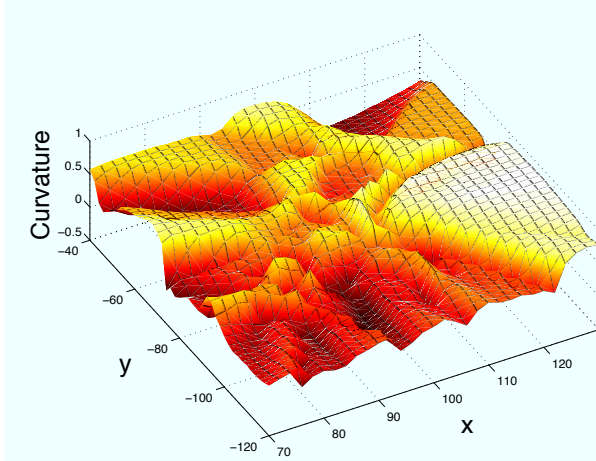
of charts to cover a manifold $\mathbf{X}$ employ neighboring disjoint charts [20, 19, 3, 2, 8]. Two charts, ψ_i and ψ_j are disjoint *iff* $\psi_i \bigcap \psi_j = \oslash$. This implies that there is no *apriori* condition to ensure that the mapping $\psi_j \circ \psi_i^{-1}$ is a *diffeomorphism* leading to high-subspace angles between the neighboring charts. A variety of nonlinear regularization techniques have been proposed to minimize the high-subspace angles [15, 21, 4]. However, by simply ensuring that there is data-overlap between the charts, $\psi_i \bigcap \psi_j \neq \oslash$, we can compute neighboring charts that have low-subspace angles between them (see Figure 4).

While a collection of local linear charts can model the motion manifold $\dot{\mathbf{X}}$, it cannot be used directly to regularize the estimates of motion. To estimate human motion at runtime, we need to chose a chart $\mathbf{V}_j$ that is an "appropriate" model for motion estimation. However, since there is no method to localize on the motion manifold to chose the appropriate model, an exhaustive search among all charts is the only option, which is computationally infeasible.
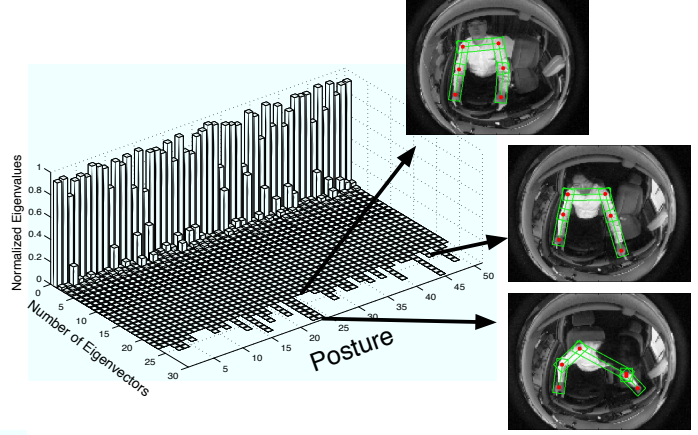
In this paper, we propose to instead exploit the dependency that exist between posture and motion to solve the model selection problem on the motion manifold. Given a set of postures configurations $\mathbf{X}$ and their associated motion configurations $\dot{\mathbf{X}}$, where $\mathbf{x} \in \mathbf{X}$ and $\dot{\mathbf{x}} \in \dot{\mathbf{X}}$ form an ordered posture-motion pair $(\mathbf{x}, \dot{\mathbf{x}}) \in \mathbf{X} \times \dot{\mathbf{X}}$ in $\mathbb{R}^{2P}$, we learn a collection of *anchored* charts for the motion manifold. The key insight here is that given a particular posture, the variation in motion around that posture can be well-approximated by a local linear model. A local linear model is a set of basis encoding the chart $\psi(\dot{\mathbf{x}})$ on the motion manifold, where $\dot{\mathbf{x}}$ is the associated motion configuration for $\mathbf{x}$. Taken together, the posture and the motion manifold are known as a *product manifold* [12].

Definition Let $\mathbf{X}$ and $\dot{\mathbf{X}}$ be covered by atlas $\{(\mathbf{U}_i, \phi_i)\}$ and $\{(\mathbf{V}_j, \psi_j)\}$ respectively. Then the sets $\mathbf{U}_i \times \mathbf{V}_j$ cover the product manifold $\mathbf{X} \times \dot{\mathbf{X}}$, and the mapping $\phi \times \psi : \mathbf{U}_i \times \mathbf{V}_j \to \mathbb{R}^P \times \mathbb{R}^P$, defined by

$$(\phi_i \times \psi_j)(\mathbf{x}, \dot{\mathbf{x}}) = (\phi_i(\mathbf{x}), \psi_j(\dot{\mathbf{x}})),$$

(a)	(b)

Figure 6. (a) Curvature of the irregular motion manifold projected into three dimensions. High curvature represents highly structured human activities while low curvature represents large amounts of motion variation around posture configurations. (b) Dimensionality of different local linear models. The arrows represent cluster center used for learning the local linear models. One can note that different postures have different local dimensionality required to capture the same amount of variance. Each Local Linear Model (LLM) can have a different dimensionality and a collection of LLMs, thus, does not require commitment to a unique latent dimensionality across the manifold.
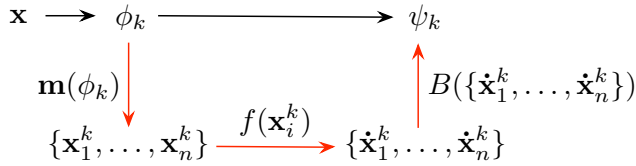

Figure 5. Red arrows show the training procedure to associate the posture charts ϕ_k with the motion charts ψ_k. Black arrows show the run-time procedure to estimate motion for posture $\mathbf{x}$ using the motion chart ψ_k. **Figure best viewed in color.**

is an injection of $\mathbf{U}_i \times \mathbf{V}_j$ onto an open subset of $\mathbb{R}^{2P}$.

By encoding the posture and the motion manifold as a product manifold we can, thus, define a collection of Local Linear Models (LLMs) that encode a set of basis for each motion chart $\psi_\mathbf{i}(\dot{\mathbf{x}})$ corresponding to a posture chart $\phi_\mathbf{i}(\mathbf{x})$ anchored at the posture configuration $\mathbf{x}$. Figure 2 shows a collection of LLMs ψ_i anchored at posture configuration $\mathbf{x}_i$'s that explains the motion variation in the charts $\phi(\mathbf{x_i})$.

Figure 5 (red arrows) shows the procedure to compute a collection of ψ_i for the posture and the motion manifold. Given a set of posture configurations $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, we first chose a collection of charts ϕ_k anchored at posture configurations $\mathbf{x}'_k$ (chosen randomly or using k-means) to cover the posture manifold. Given a chart ϕ_k, the membership function $\mathbf{m}(\phi_\mathbf{k})$ return a set of n nearest posture neighbors using the Euclidean distance metric. Associated with every posture $\mathbf{x}_j^k \in \mathbf{m}(\phi_\mathbf{k})$ is a motion configuration $\dot{\mathbf{x}}_j$. The set $\dot{\mathbf{X}}_i^{NN} = [\dot{\mathbf{x}}_1^k, \ldots, \dot{\mathbf{x}}_n^k]$ is then a representative sample of motion variation around the posture configuration $\mathbf{x}'_i$.

A local linear model ψ_k is computed using Principal Component Analysis (PCA) and the requisite number of basis vectors $B(\{\dot{\mathbf{x}}_1^k, \ldots, \dot{\mathbf{x}}_n^k\})$ that sum up to a pre-defined energy level are retained.

Figure 6 (a) shows the curvature at different parts of the motion manifold. It is evident that there is a lot of variance in the local curvature at different parts of the motion manifold. Each local linear model can have a different dimensionality, and taken together as a collection of local linear models, serve as an efficient representation of the irregular motion manifold. This avoids the limitation of global models that require commitment to a uniform dimensionality across the motion manifold [11, 22]. Figure 6 (b) shows the dimensionality for the local linear models anchored at different posture configurations. It is evident that the requisite number of basis to capture a pre-defined energy level is different for different posture configurations, motion around some postures is highly structured and thus requiring less number of eigenvectors and vice-versa. A collection of local linear models, thereby, allows an efficient and compact representation of the irregular motion manifold.

4. Motion Estimation over the Product Manifold

We have used the collection of local linear models to design a subspace-driven motion estimation algorithm. At runtime, Figure 5 (black arrows) shows how ψ_i can be used to estimate motion. Given a posture configuration $\mathbf{x}$, we find the closest posture chart ϕ_k as defined by the distance to its posture anchor $\mathbf{x}'_k$. Associated with the chosen posture

chart ϕ_k is the motion chart ψ_k that is the appropriate local linear model for motion estimation for $\mathbf{x}$, thereby, solving the model selection problem for the motion manifold. A favorable characteristic of our formulation is that the subspace constraints are enforced during the motion estimation process and not "after the fact".

Articulated motion estimation can be posed as a constrained linear-least squares problem combining the image brightness constancy constraints with the articulation constraints [5],

$$\min_{\mathcal{A}} \|\mathcal{B} - \boldsymbol{\Gamma}\mathcal{A}\|_2^2 \ \text{subject to} \ \boldsymbol{\Theta}\mathcal{A} = 0, \qquad (1)$$

where $\mathcal{A}$ are the affine parameters, $\boldsymbol{\Gamma}$ is an $M \times N$ matrix of the brightness constancy constraints, $\mathcal{B}$ is a M-vector, $\boldsymbol{\Theta}$ is a $C \times N$ matrix of the articulation constraints.

The relationship between the unconstrained affine parameters $\mathcal{A}$ in Equation 1 and the subspace-constrained affine parameters $\mathcal{A}^*$ can be expressed as,

$$\mathcal{A} = \mathcal{S}_P \mathcal{A}^* + \bar{\mathcal{A}}, \qquad (2)$$

where $\mathcal{S}_P$ is the local linear model anchored at the posture configuration P and $\bar{\mathcal{A}}$ is the mean motion vector. We can now derive the subspace-constrained formulation for motion estimation from Equation 1,

$$\min_{\mathcal{A}^*} \|\mathcal{B} - \boldsymbol{\Gamma}(\mathcal{S}_P\mathcal{A}^* + \bar{\mathcal{A}})\|_2^2 \ \text{subject to} \ \boldsymbol{\Theta}(\mathcal{S}_P\mathcal{A}^* + \bar{\mathcal{A}}) = 0, \qquad (3)$$

Equation 3 formulates motion estimation as a nonlinear minimization problem between the image brightness constancy and the articulation constraints within a particular local linear model. Upon some algebraic substitution we can express Equation 3 as,

$$\min_{\mathcal{A}^*} \|\tilde{\mathcal{B}} - \tilde{\boldsymbol{\Gamma}}\mathcal{A}^*\|_2^2 \ \text{subject to} \ \tilde{\boldsymbol{\Theta}}\mathcal{A}^* = -\hat{\boldsymbol{\Theta}}. \qquad (4)$$

Equation 4 is an exact-equality constrained linear least-square system that be solved using the Lagrange Multipliers,

$$f(\mathcal{A}^*|\lambda) = \|\tilde{\mathcal{B}} - \tilde{\boldsymbol{\Gamma}}\mathcal{A}^*\|_2^2 + 2\lambda(\tilde{\boldsymbol{\Theta}}\mathcal{A}^* + \hat{\boldsymbol{\Theta}}). \qquad (5)$$

The gradient of $f(\mathcal{A}^*|\lambda)$ equals zero when,

$$\tilde{\boldsymbol{\Gamma}}^T\tilde{\boldsymbol{\Gamma}}\mathcal{A}^* + \tilde{\boldsymbol{\Theta}}^T\lambda = \tilde{\boldsymbol{\Gamma}}^T\tilde{\mathcal{B}}, \qquad (6)$$

and

$$\tilde{\boldsymbol{\Theta}}\mathcal{A}^* = -\hat{\boldsymbol{\Theta}}. \qquad (7)$$

This can be written and solved as a Karush-Kuhn-Tucker system [9],

$$\begin{bmatrix} \tilde{\boldsymbol{\Gamma}}^T\tilde{\boldsymbol{\Gamma}} & \tilde{\boldsymbol{\Theta}}^T \\ \tilde{\boldsymbol{\Theta}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathcal{A}^* \\ \lambda \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{\Gamma}}^T\tilde{\mathcal{B}} \\ -\hat{\boldsymbol{\Theta}} \end{bmatrix}. \qquad (8)$$

To update the posture configuration for the next time-step, we can re-project the subspace-constrained affine motion parameters $\mathcal{A}^*$ into the original 30 dimensional motion configuration space using Equation 2.

5. Results

We have conducted several experiments to showcase the benefits of the proposed approach. In particular, we compute motion estimates using the proposed approach and compare it against the tracking method proposed by Datta *et al.* in [5]. In addition, we compare the proposed approach against global models of the motion manifold obtained using the bilinear model [18] and PCA. A Matlab implementation of the motion estimation algorithm over the motion manifold ran at up to 50 frames per second (fps).

A labeled data set of two humans performing several activities during the process of driving a car was collected. Each actor was captured for approximately 3 minutes at 60 fps at an image resolution of 160×120. The data set consists of approximately 18,000 frames. This labeled data set was then used to learn a collection of local linear models for long-term human motion estimation. We used 500 clusters with 400 nearest neighbors to learn a collection of local linear models (LLMs) using weighted PCA.

We manually initialize six points in the first frame of the image sequence. Our human upper body model has 6 points encoding the posture configuration with 4 articulation constraints for the two elbows and shoulders. Figure 8 presents the motion estimation result using our implementation of [5], PCA, and the proposed collection of local linear models respectively. It can be observed that both [5] and PCA built up an appreciable amount of drift and lost track of the right shoulder as opposed to the collection of local linear models. Figure 9 shows an image sequence of an actor reaching out for the rear-view mirror of a car. It can be observed that both the method of [5] and PCA fail to track the actor. The collection of local linear models correctly tracks the actor, thereby highlighting the benefit of the proposed approach. Figure 10 shows motion estimation for an image sequence that involves an actor performing a series of walk, jump, walk action. Figure 10 (top row) shows the motion estimates obtained by [5]. It is evident that that [5] has lost track of the left elbow. We collected a data set of 500 labeled frames to learn a collection of local linear models for this set of human actions. Figure 10 (bottom row) shows the motion estimates from the collection of local linear models and it can be observed that the recovered tracks are correct.

In another experiment, we collected a labeled data set of 2,000 frames of a person performing various activities in a desktop environment such as using the mouse, typing on the keyboard, touching face, among others. Figure 11 (top row) shows the computed motion using the algorithm from [5]. It can be observed that [5] builds an appreciable amount of drift and loses tracks of the shoulders. The collection of local linear models, on the other hand, maintain correct tracks over the course of the image sequence (bottom row of Figure 11).

5.1. Comparison to Bilinear Models

Tenenbaum and Freeman in [18] introduced the bilinear model as a global model for a manifold. Bilinear models are more expressive than the linear models because the two factors can modulate the model multiplicatively. For more details, we refer the interested reader to [18]. Figure 12 compares motion estimation using a collection of LLMs against the Bilinear model. It can be observed that the LLMs produce accurate tracks. This and earlier results (see Figure 8 and 9) demonstrate that global models (bilinear or linear) are ineffectual for modeling the irregular motion manifold.

Finally, Figure 7 shows a quantitative comparison between the proposed collection of LLMs, PCA motion model, Bilinear motion model and the approach of Datta *et al.* [5]. Figure 7 (a) shows the average Root-Mean-Square (RMS) error for results shown in Figure 8 and Figure 9. The large increase in error in the previous approaches is due to the failure to track the right arm in Figure 9. Figure 7 (b) shows the average RMS error for Figures 10 and 11. The previous approaches fail to track the left elbow in the walk-jump-walk sequence (Figure 10) and accumulate drift during desktop tracking sequence (Figure 11). Figure 7 (c) shows that the previous approaches drift during tracking of white shirt with very little texture (Figure 12).

5.2. Anomalous Action Detection

Since we learn a model of long-term human motion, we can use it detect instances when the computed motion starts to differ from the learnt model. Figure 13 shows one such application for novel action detection. In this case, the action of a human exiting the car was not included in the database of labeled human posture configurations. We can observe that the method of [5] has no appreciable increase in the associated uncertainty, illustrated using the covariance ellipses, when the person starts to exit the vehicle. Our method, on the other hand, demonstrates appreciable increase in posture uncertainty. Such an application could have use in safety applications or human machine interface.

6. Conclusions

In this paper, we have proposed an approach that exploits the dependencies between posture and motion for human motion estimation. Given a particular posture configuration, the motion around that posture is highly structured and can be well-approximated by a local linear model. Long-term human motion consists of human activities in an arbitrary order and a collection of local linear models can compactly represent the unordered dynamics of human motion. Each local linear model can have a different dimensionality and taken together as a collection, local linear models avoid the limitations of the traditional approaches that require a uniform latent dimensionality across the motion manifold.

We have used the collection of local linear models to design a subspace-constrained articulated motion estimation algorithm that produces accurate and longer tracks compared to the literature [5]. Moreover, we compared our motion estimation against PCA and the bilinear model [18] and found that a global model for the irregular motion manifold is ineffectual. In future work, we hope to extend a collection of local linear models from currently encoding frame-to-frame motion to encoding motion between groups of frames for higher level modeling of human motion.

References

[1] C. B. Abdelkader. Motion-based recognition of people in eigengait space. In *FGR*, 2002.

[2] A. Agarwal and B. Triggs. Tracking articulated motion with piecewise learned dynamic models. *ECCV*, 2004.

[3] C. M. Bishop and C. K. I. Williams. Gtm: The generative topographic mapping. *Neural Computation*, 1998.

[4] M. Brand. Charting a manifold. In *NIPS*, 2003.

[5] A. Datta, Y. A. Sheikh, and T. Kanade. Linear motion estimation for systems of articulated planes. In *CVPR*, June 2008.

[6] J. Duetscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. *CVPR*, 2, 2000.

[7] M. Isard and A. Blake. Condensation conditional density propagation for visual tracking. *IJCV*, 29, 1998.

[8] N. Kambhatla and T. K. Leen. Dimension reduction by local principal component analysis. *Neural Computation*, 1997.

[9] H. Kuhn and A. Tucker. Nonlinear programming. In *Second Berkeley Symposium on Mathematical Statistics and Probability*, 1951.

[10] N. D. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *NIPS*, 2004.

[11] C.-S. Lee and A. M. Elgammal. Modeling view and posture manifolds for tracking. In *ICCV*, 2007.

[12] D. Martin. *Manifold Theory: An Introduction for Mathematical Physicists*. Albion/Horwood Publishing Ltd, 2002.

[13] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *PRL*, 1996.

[14] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 1901.

[15] S. Roweis and G. E. Hinton. Global coordination of local linear models. In *NIPS*, 2002.

[16] H. Sidenbladh, M. J. Black, and D. J. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *ECCV*, 2000.

[17] L. M. Tanco and A. Hilton. Realistic synthesis of novel human movements from a database of motion capture examples. In *Workshop on Human Motion*, 2000.

[18] J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 2000.

[19] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 1999.

[20] R. Urtasun, D. J. Fleet, and P. Fua. Temporal motion models for monocular and multiview 3d human body tracking. *CVIU*, 2006.

[21] J. J. Verbeek, S. T. Roweis, and N. Vlassis. Non-linear cca and pca by alignment of local models. In *NIPS*, 2004.

[22] J. M. Wang, D. J. Fleet, and A. Hertzmann. Gaussian process dynamical models. In *NIPS*, 2006.
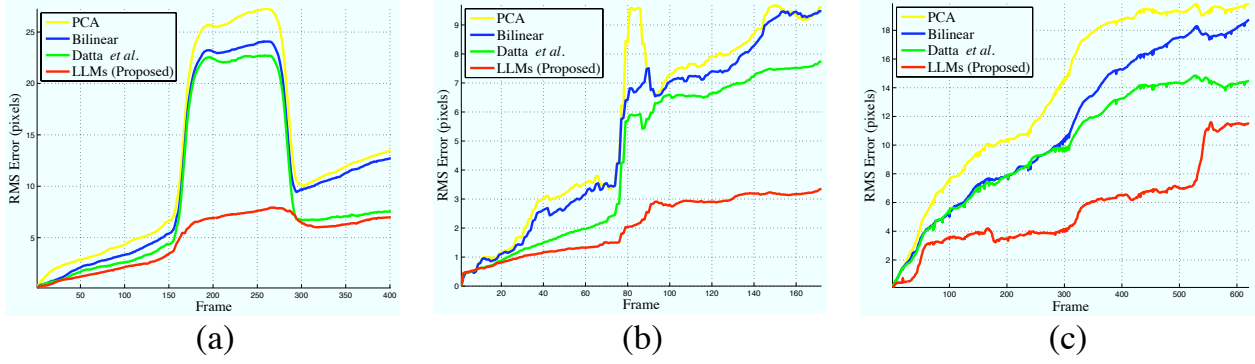
Figure 7. Motion estimation comparison between the proposed collection of Local Linear Models (LLMs) and PCA motion model, Bilinear motion model, and Datta *et al.* [5]. Average RMS motion estimation error is shown in (a) for Figure 8 and Figure 9, in (b) for Figure 10 and Figure 11 and in (c) for Figure 12.
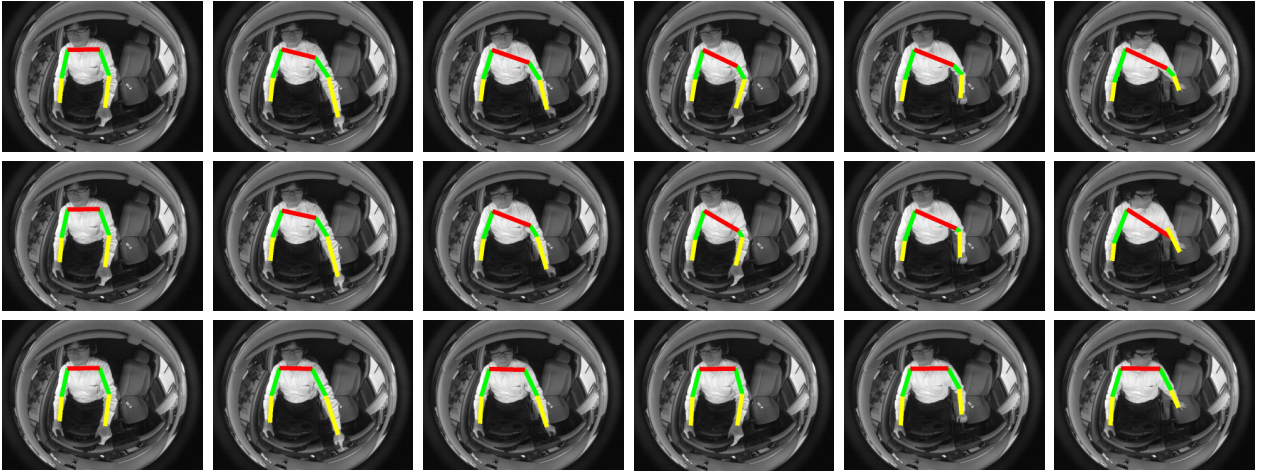

Figure 8. Top row: motion estimation using the method describe in [5]. Middle row: motion estimation using a global Principal Component Analysis (PCA) model of the motion manifold. Bottom row: motion estimation using the proposed collection of local linear models. It can be observed that the proposed method outperforms both [5] and the PCA model of the motion manifold.
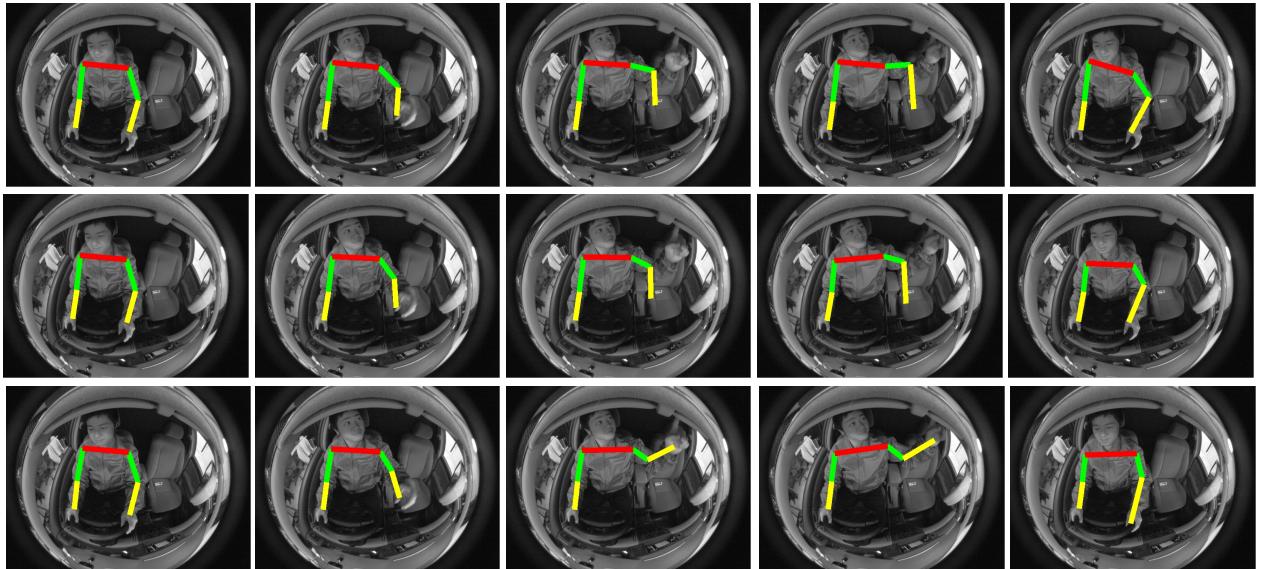
Figure 9. Top row: motion estimation using [5]. Middle row: motion estimation using global PCA model. Bottom row: motion estimation using a collection of local linear models. It is evident that the proposed method correctly tracks the actor, while [5] and PCA do not.
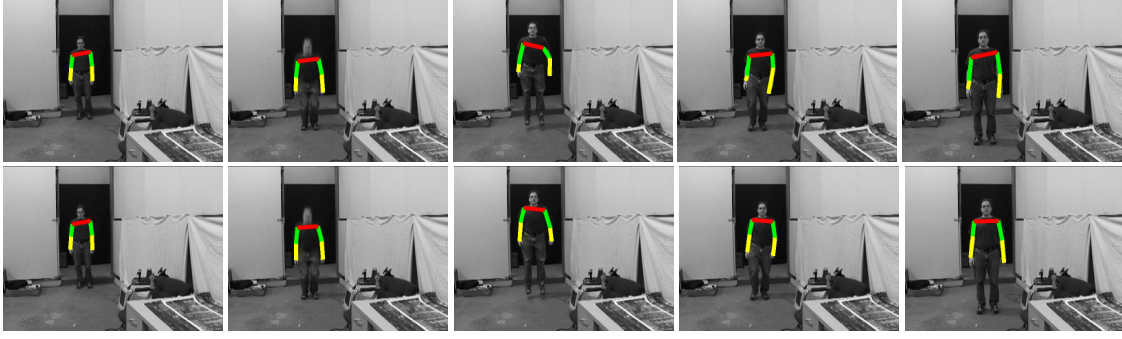
Figure 10. Motion estimation comparison for a walk-jump-walk image sequence. Top row: motion estimation using the method describe in [5]. Bottom row: motion estimation using the proposed collection of local linear models. Note the incorrect elbow location in the right-most column for tracks obtained from [5].
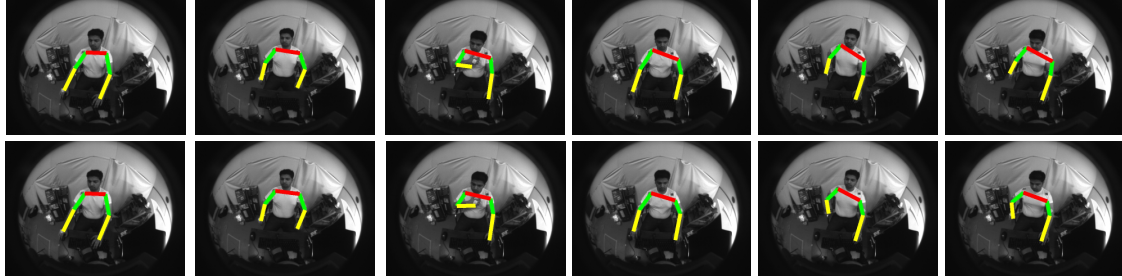

Figure 11. Collection of local linear models are applicable in different environments. In this example, we show tracking in a Desktop environment. Top row: motion estimation using [5]. Bottom row: motion estimation using the proposed collection of local linear models. Note the incorrect posture configuration in the right-most column for tracks obtained from [5].
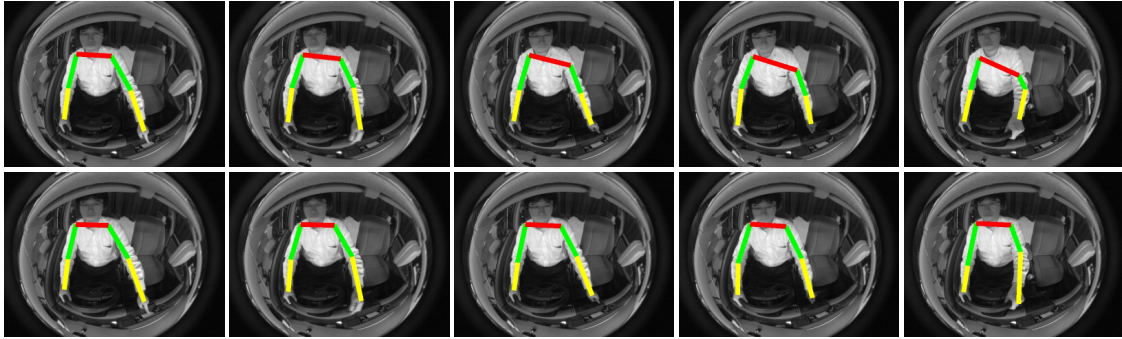

Figure 12. Comparison with the bilinear model [18]. Top row: Motion estimation result using the bilinear model. Bottom row: Motion estimation using the proposed collection of local linear models. It is evident that the proposed method outperforms the bilinear model.
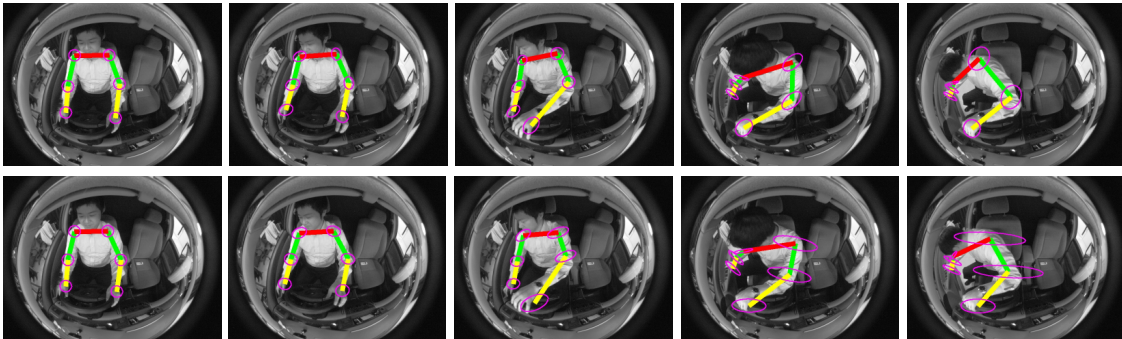

Figure 13. Anomalous action detection. Top row: posture configuration covariance ellipses from [5]. Bottom row: posture configuration covariance ellipses using the proposed collection of local linear models. The covariance ellipses grow rapidly if the actor performs a novel action helping to detect anomalous actions.