

Automatic configuration of spectral dimensionality reduction methods for 3D human pose estimation

Michał Lewandowski
DIRC Kingston University
London, UK, KT1 2EE
m.lewandowski@kingston.ac.uk

Dimitrios Makris
DIRC Kingston University
London, UK, KT1 2EE
d.makris@kingston.ac.uk

Jean-Christophe Nebel
DIRC Kingston University
London, UK, KT1 2EE
j.nebel@kingston.ac.uk

Abstract

In this paper, our main contribution is a framework for the automatic configuration of any spectral dimensionality reduction methods. This is achieved, first, by introducing the mutual information measure to assess the quality of discovered embedded spaces. Secondly, we overcome the deficiency of mapping function in spectral dimensionality reduction approaches by proposing data projection between spaces based on fully automatic and dynamically adjustable Radial Basis Function network. Finally, this automatic framework is evaluated in the context of 3D human pose estimation. We demonstrate mutual information measure outperforms all current space assessment metrics. Moreover, experiments show the mapping associated to the induced embedded space displays good generalization properties. In particular, it allows improvement of accuracy by around 30% when refining 3D pose estimates of a walking sequence produced by an activity independent method.

1. Introduction

3D pose recovery from video sequences is a very active research field in computer vision, which has many applications including visual surveillance, sport analysis and human computer interaction. Although this very challenging task has been tackled by many groups for years [1; 2; 3; 4; 5; 6; 7; 8; 9], the quality of recovered 3D poses still requires improving. The objective of this paper is the automatic configuration of spectral dimensionality reduction methods applied to the refinement of pose estimates extracted from video sequences.

Here, 3D motions are defined as sequences of 3D human body poses at successive time instants, thus, motion reconstruction can be formulated as the problem of reconstruction the sequence of 3D human poses. In turn, the goal of pose recovery is to localise a person's joints and limbs in either an image plan (2D recovery) or a world space (3D recovery), which usually results in the reconstruction of a human skeleton. Although, human motions reside in a very high dimensional parameter space;

many studies have revealed that the space of a specific activity, such as walking, is intrinsically a low dimensional nonlinear subspace embedded in the high dimensional space [1; 2; 7]. Therefore, reduction of data dimensionality has been investigated to improve 3D pose recovery.

In this work, we assume that some ambiguous 3D poses have been estimated from 2D images cues and we concentrate on refinement the quality of these poses by taking advantage of advanced nonlinear activity-constrained learning based approaches. The proposed method relies on discovering the embedded space associated to the activity of interest in fully automatic way without manual tuning of any parameter. It is invariant to body size or motion style, moreover the accuracy increase with the size of training set. Since it is strictly activity specific, training and testing data must originate from the same domain of motion (e.g. walking). The discussed method is general and can be used either for improving 2D or 3D pose estimation depending on training set dimensionality.

The paper contributes to the current state of the art, first, by providing a general framework for the automatic configuration of any spectral dimensionality reduction methods. This is achieved by introducing the mutual information measure [10] to assess the quality of discovered embedding spaces. Secondly, we overcome the deficiency of mapping function in spectral dimensionality reduction approaches by exploiting Radial Basis Function network (RBFN) [11]. In our method, the RBFN can dynamically adjust to the training data to design the optimal structure in an unsupervised manner. This automatic framework is evaluated in the context of refining of 3D human pose estimations obtained by any activity independent 3D pose recovery method. In particular, we identify the most appropriate spectral dimensionality reduction method for human motion. Finally, we demonstrate that mutual information outperforms all currently available metrics for judging the quality of the embedded space by providing comprehensive and quantitative comparison between them.

The structure of this paper is organised as follows. After presenting relevant literature review in Section 2, we detail in Section 3 our automatic configuration algorithm and explain its application. Then, experiments to validate our

method with quantitative results are given in Section 4. Eventually, conclusions and future work are addressed in Section 5.

2. Previous work

2.1. 3D pose reconstruction.

The inference of 3D pose is derived from information obtained during image processing. Once these informative cues have been extracted, 3D pose reconstruction methods fall into one of two categories: activity independent [8; 12; 6; 13; 5; 4] and activity-constrained learning [1; 2; 14; 3; 7; 9] approaches.

Although, there is a great variety of different activity independent approaches, most of them presuppose an explicitly known parametric body model and require a good initialization. Some methods estimate the pose by exploiting simple orthographic projection model like in [8; 12]; whereas others take the advantage of the pin-hole camera model and auto calibration technique [4]. A data-driven iterative approach is presented in [5], where pose candidates are generated in Markov chain Monte Carlo search guided by image observations. Eventually, where multiple views are available the pose can be inferred by visual-hull reconstruction algorithm [6; 13].

Activity-constrained learning approaches focus on learning the prior model of motion directly from carefully selected training data provided by motion capture systems (MoCap). Example based approaches [3] explicitly store training images of silhouettes from various viewpoints with corresponding 3D poses so the entire space of possible solutions is covered. The final pose is inferred by interpolating between poses which are obtained by searching the closest matching features to a given input query. A main limitation of these techniques is that a very large training set is required to provide satisfactory accuracy and generalization properties. In contrast, dimensionality reduction techniques [1; 2; 14; 7; 9] learn a low dimensional informative and compact representation of training set which can be used to recover 3D poses with the usage of mapping functions provided directly by the dimensionality reduction method [2; 7] or learned in an additional step [1; 14].

Since human motion is not linear, nonlinear dimensionality reduction techniques are required to provide complex embeddings of high dimensional data. Mapping based approaches as GPLVM [2; 9] focus on learning the probabilistic nonlinear mapping functions during dimensionality reduction, whereas embedded based approaches (so-called spectral methods) [15; 16; 17] discover directly the structure of the underlying manifold. Since the learning process of mapping based approaches is computationally very expensive, their practical application results of using small training datasets which may

compromise accuracy. On the other hand, in spectral methods the accuracy of estimation may scale proportionally with the size of training set with a reasonable computational cost when an appropriate mapping is provided.

2.2. Dimensionality reduction techniques based on spectral approaches

Spectral or embedding-based approaches model the structure of data by preserving some geometrical property of the underlying manifold. While isometric feature mapping (Isomap) method [16] attempts to maintain global properties, Laplacian Eigenmaps (LE) [17] and Locally Linear Embedding (LLE) [15] methods aim at preserving local geometry which implicitly tends to preserve the global layout of the data manifold.

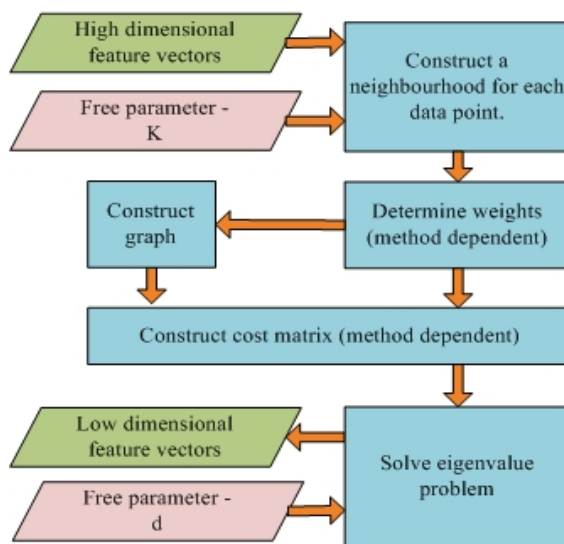


Figure 1: Process of dimensionality reduction using spectral methods.

These methods share the same structure of algorithm which is illustrated in Figure 1. First, the neighbourhood for each data point is constructed by choosing K -nearest neighbours based on Euclidean distances. Then, weights, which express the geometrical relationship between each data point and its neighbours, are determined according to the property to be preserved. In LLE, they summarize the contribution in reconstruction of a data point from its neighbours [15]. In LE, they approximate the distance between a point and its neighbours [17]. Finally, in Isomap, weights correspond to Euclidean distances between a point and its neighbours [16]. Subsequently, LE and Isomap algorithms approximate the manifold by an adjacency graph where nodes correspond to data point and edges represent weights between points. Then, each method optimises its own cost function subject constraints that make the problem well-posed: a sparse cost matrix is constructed directly for LLE and LE [15; 17], whereas the

Isomap dense cost matrix is obtained by calculating geodesic distances between all pairs of points in the graph [16]. Finally, spectral embedding is calculated using the eigenvectors of the cost matrix.

All approaches have two free parameters: ‘d’ and ‘K’. ‘d’ is the dimensionality of the embedded space and must be known a priori because it is used in the minimization process. If the number of dimensions is too low, important data features may be collapsed onto the same dimension. Similarly, ‘K’, i.e. the neighbourhood size, cannot be too small since global feature information could be lost if the manifold is split into unconnected pieces. Moreover, the number of neighbours cannot be too large because local methods rely on the assumption that a data point and its nearest neighbours can be modelled as locally linear. In the case of Isomap, choosing K too large, may result in including data points from other branches of the manifold and introducing an error in geodesic distances.

Since the effectiveness of a method depends on the choice of these parameters, automatic selection techniques have been proposed. The best choice for ‘d’, i.e. the estimation of the intrinsic dimension of the high dimensional data, has been well studied in the literature. Many techniques have been suggested: maximal likelihood estimation (MLE) [18], packing numbers [19] or analysis of a geodesic minimum spanning tree [20]. However, since this problem has been solved satisfactory by MLE, this method is generally used for estimation of intrinsic dimensionality.

Some quantitative measures have also been proposed for estimation of the optimal value of ‘K’. This is achieved by assessing directly the quality of embedded spaces by different metrics. They include Residual Variance, Spearman Rho and Procrustes Measure. The residual variance [21; 22] expresses how well the distance information is preserved between two sets of variables, i.e. it reflects the degree of linear relationship between these variables. Spearman’s rho [23] measures the accuracy of the low-dimensional manifold in retaining the order of pair wise distances of data points of the high-dimensional. Finally, the procrustes measure [24] reflects the matching of two sets of variables in terms of distances. It determines

how well a linear transformation (translation, reflection, orthogonal rotation, and scaling) of the points in one space conforms to the points in the second space. Experiments have revealed all these measures are data dependent and therefore cannot be trusted for the automatic selection of free parameter ‘K’ in an untested domain. In this work, we deal with this fundamental issue by adopting a new metric, called mutual information, which outperforms the other measures and allows a much better automatic selection of ‘K’, which is essential in our framework.

3. Automatic configuration of spectral dimensionality reduction methods

To demonstrate the performance of the proposed methodology for automatic configuration of spectral dimensionality reduction methods, it is applied to the refinement of 3D body pose estimates. The framework presented in Figure 2 is composed of two parts: activity learning, which is an automatic offline process, and the online procedure of pose refinement.

During the learning stage, MLE is used to evaluate the intrinsic dimensionality ‘d’ of the activity of interest represented by 3D motion capture data $X = \{x_i\}_{i=1..N}$ (MoCap). Then, a spectral method is applied on sequence of ‘D’ dimensional MoCap data to reduce dimensionality to ‘d’ ($d < D$). The reduction of dimensionality is repeated for a range of ‘K’ values. For each ‘K’ value, the quality of the discovered embedded space is evaluated using a quantitative metric. In the next step, the best low dimensional space according to the metric is chosen ($Y = \{y_i\}_{i=1..N}$) and used first for finding the structure of the RBFN and subsequently for training the network.

The online module of the framework solves the actual problem of 3D pose recovery. In principle, it can be applied to pose estimates produced by any activity independent methods [8; 12; 25; 26; 4; 13; 6]. In this paper, we consider the output of Kuo’s algorithm as the sequence of 3D pose estimates [4]. Each pose is projected to the selected low dimensional space using RBFN [11], then refined and projected back to high dimensional space in order to produce the final pose estimation.

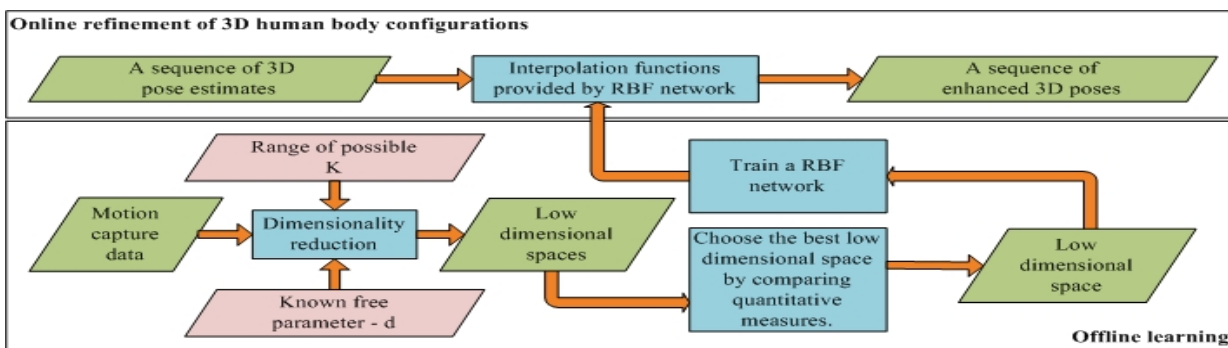


Figure 2: Automatic refinement of 3D body pose estimates.

3.1. Automatic training of RBF network

All spectral approaches suffer from the deficiency of not providing a mapping function. We address it by introducing unsupervised Radial Basis Function network (RBFN). Although these networks have already been used in similar scenario [14; 1], our contribution is that their structure dynamically adjust to the training set rather than be set manually. RBFN is expressed by:

$$y = f(x) = B \psi(x) \quad (1)$$

where ‘B’ is the matrix of network weights and ψ is a vector of radial basis functions $[\Phi(u_1), \dots, \Phi(u_{N_c})]$. We exploit Gaussian basis function ($\Phi(u) = e^{-au}$) in place of a thin plate spline basis function [1; 14] which has excellent approximation properties [11]. ‘u’ denotes the Euclidean distance $\|x - c_i\|$ between a point and the i^{th} cluster centre, whereas $a = 0.5\sigma^{-2}$ is a scaling parameter. The RBFN structure is formed by centres ‘c’ which summarize training data points to provide good generalization properties. Since the performance of RBFN critically depends upon the chosen centres [27], we incorporate into the framework the rival penalized competitive learning technique proposed in [28], which contrarily to k-means clustering algorithm [29] is capable to find the optimal localization of centres as well as their correct number N_c in an automatic way. The last parameter ‘ σ ’ is the width of the clusters which is taken as the average distance between all centres. When the centres and their spreads are determined, then the network weights ‘B’ are learned by solving least square problem for over constrained nonlinear system of equations: $Y = B \psi(X)$.

In our approach, the two mapping functions are being learned separately, i.e. from low dimensional to high dimensional space (f_{dD}) and in the opposite direction (f_{Da}), so that parameters can be optimized independently. The mapping process consists of the following steps. Given some unseen input x , first, the pose is projected into low dimensional representation y :

$$y = f_{Da}(x) \quad (2)$$

$$f_{Da}(x) = B_{Da} * [\Phi(\|x - c_{1Da}\|), \dots, \Phi(\|x - c_{NcDa}\|)] \quad (3)$$

In low dimensional space we search for the closest two points, y_1 and y_2 , in the manifold and interpolate between them to calculate y' , which should correspond to the most similar pose in 3D space:

$$y_1 = \min \|y - y_i\|, i = 1..N \quad (4)$$

$$y_2 = \min \|y - y_j\|, j = 1..N \text{ and } j \neq i \quad (5)$$

Finally, we project y' to original space to generate the enhanced 3D pose, x' :

$$x' = f_{dD}(y') \quad (6)$$

$$f_{dD}(y') = B_{dD} * [\Phi(\|y' - c_{1dD}\|), \dots, \Phi(\|y' - c_{NcdD}\|)] \quad (7)$$

3.2. Automatic selection of neighbourhood size

Determination of the best value of neighbourhood size ‘K’ is performed by judging quality of an embedded space by the holdout method. This is achieved by separating data into two sets: training and testing. First, the dimensionality of training set is reduced. Then, the RBFNs are trained using the produced embedding with original spaces. Since projection and refinement of testing data between spaces results in a set of pose predictions, their errors can be calculated to evaluate different values of ‘K’. The optimal neighbourhood size is the value of ‘K’ that can be used to produce the embedded space with corresponding smallest error. In our approach, we would like to identify the best embedded space in an unsupervised manner without calculating computational expensive prediction errors for all possible values of ‘K’.

We propose using mutual information [10] (MI) to assess the quality of discovered embedded spaces which has never been used in this context. Since MI is able to discover even marginal dependency between two spaces of variables, it can be applied to measure dependency between low and high dimensional spaces, respectively Y and X. Therefore, the best value of ‘K’ should provide the maximum value of MI for a given dimensionality reduction method. MI is expressed by the following formula [10]:

$$I(Y, X) = \sum_{x \in X} \sum_{y \in Y} p(y, x) \log \left(\frac{p(y, x)}{p(y)p(x)} \right) \quad (8)$$

Nevertheless, since the estimation of MI based on entropy is only applicable in one dimension, we use k-nearest neighbour statistics for determining MI in higher dimensions as implemented by [10].

4. Experimental results

4.1. Datasets

The proposed framework was tested on the HumanEva (HE) dataset [30], whose video and Mocap data are used as benchmark for pose recovery. They also provide a standard set of error metrics to evaluate pose estimations [31]. Before using MoCap data, they need to be converted into *normalized poses* [14], i.e. invariant to subject’s rotation and translation. Then, the three angles defining each joint position is represented by a quaternion. Since each 3D pose is described originally by an articulated skeleton composed of 13 joints (3 joints per limb plus the head), each data point is expressed by a 52 dimension vector.

Sequences of “walking in a circle” provided by HE were

used for our experiments. Trail 3 of subject 3 was used for training (S3T3). We chose frames 750 to 1750 to include a variety of walking postures. Testing was performed using three ground truth datasets and one dataset composed of pose estimates. Datasets were carefully selected to validate robustness of the framework with different actors, who differ in size, body shape, motion style and gender. The ground truth datasets consist of: frames 55 to 315 for male subject 3 in trail 1 (S3T1), frames 340 to 760 for male subject 2 in trial 1 (S2T1) and frames 1 to 400 for female subject 1 in trial 1 (S1T1). The last dataset is a set of body configuration estimates obtained through auto calibration technique for subject 2 (S2EST) [4].



Figure 3: Variety of HE actors using in our experiments: from left to right, female subject 1 (testing); male subject 2 (testing); and the male subject 3 (training and testing).

4.2. Experimental framework

The overall aim of our experiments is to demonstrate the effectiveness of automatic configuration of spectral dimensionality reduction methods in the context of improving pose estimates obtained through any activity independent technique. In the first place, to obtain the highest accuracy, we need to identify the best dimensionality reduction method for human body motion data. Secondly, since the efficiency of embedding-based approaches relies on the estimation of the best neighbourhood size ‘K’, we need to evaluate the accuracy of the new MI estimator against Residual Variance (RV), Spearman Rho (SR), Procrustes Measures (PM) and the optimal value (The best) calculated using projection error as defined in section 3.2. To achieve that we performed four experiments, which extensively evaluate the performance and robustness of the proposed methodology.

In the first experiment, we consider the simplest scenario, in which we train and test with the same subject, i.e. S3, using different trials. This should allow finding the lowest bound of the reconstruction error which can be obtained within our framework. In the next two experiments, we evaluate our approach using MoCap data from subjects 2 (S2T1) and 1 (S1T1). Both datasets differ considerably from training set, because walking actions are performed by very different subjects, see Figure 3. Since input data are ideal estimates, the reconstruction error should highlight differences introduced by variations of

walking styles and body frames between testing and training characters. Finally, in the fourth experiment, we take the 3D pose estimates produced by Kuo’s algorithm [4] for subject 2 on a walking sequence (S2EST) and we refine them according to our framework (Figure 2).

All above experiments were setup to test the main spectral methods, i.e. LLE, LE and ISOMAP, and PCA, which is used as a reference. First the intrinsic dimensionality of dataset S3T3, $d=2$ was determined by MLE [18]. Then, in all cases, dataset S3T3 is used for training the spectral dimensionality reduction methods. To discover, the optimal neighbourhood size, values between 5 and 35 were used to produce 2-dimension spaces. Because PCA is independent from K, it is only computed once per experiment. As the result, for each method 36 low dimensional spaces are produced which are then used for training RBF mapping functions between spaces. Finally, prediction errors are calculated and used to compare quality of embedded spaces. An example of low dimensional manifold discovered by Isomap is depicted in Figure 4.

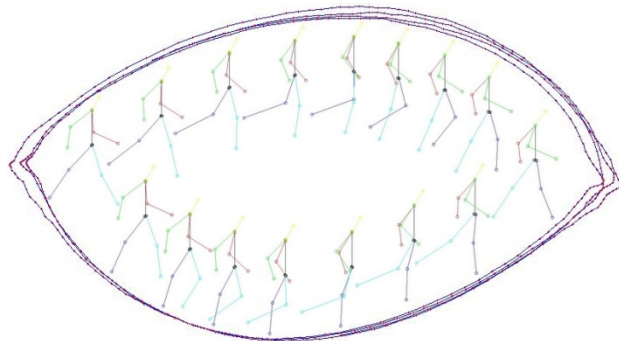


Figure 4: Representation of the best low dimensional manifold with corresponding poses discovered by Isomap according to the MI measure for training set S3T3.

Estimated poses are evaluated against motion capture data, which is our ground truth. Firstly, we report mean (over all angles) absolute difference errors between the true and estimated joint angle vectors (in degrees) to show that our method is skeleton independent:

$$MAE(^{\circ}) = \frac{1}{M} \sum_{i=1}^M |(x_i - x'_i) \bmod 180^{\circ}| \quad (9)$$

Secondly, Procrustes Analysis [32] is performed to facilitate the comparison between the reconstructed body and the ground truth data when the properties of body model are known. Procrustes Analysis determines a linear transformation (translation, rotation, and scaling) of the reconstructed body to best match to the ground truth by minimising Root-Mean-Square error (RMS).

Note that we provide, for the first time, results of practical employment of RBFN and automatic parameter selection in the context of 3D pose recovery using Laplacian Eigenmaps (LE) [13].

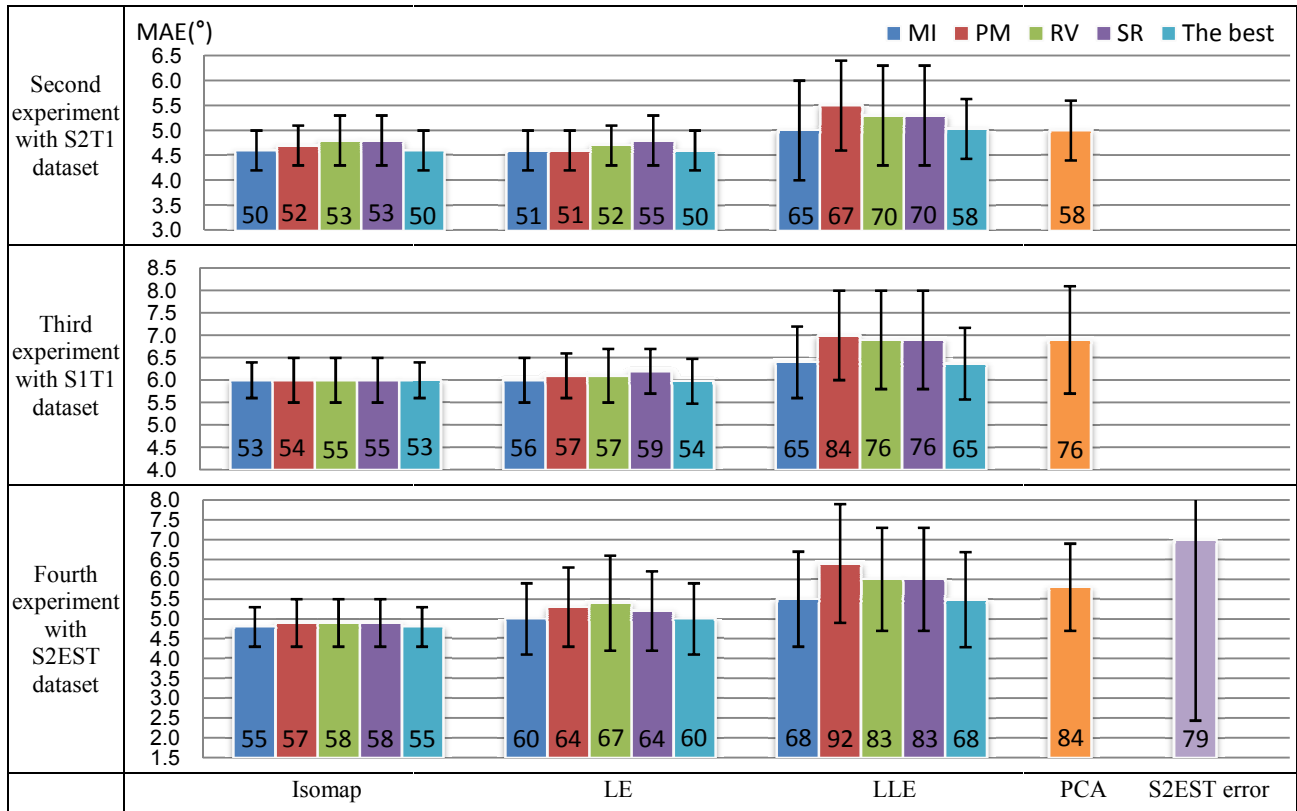


Figure 5: Mean of MAE and standard deviation for the best low dimensional spaces according to four coefficients discovered by different methods for three datasets. The RMS error in mm is depicted within bars.

4.3. Results

In the Figure 5, the MAE angle error and the corresponding RMS error are provided for all considered methods obtained in the second, third and fourth experiment. Detailed quantitative results of the first experiment are not provided, because all methods performed very well and the impact of neighbourhood size estimation on results is marginal. In the case of spectral approaches, the MAE angle error for the embedded space calculated using the optimal K ('The best') is compared with those obtained for the low dimensional space selected by the four quantitative measures.

Comparison of dimensionality reduction methods reveals that Isomap and LE are almost equally good in improving accuracy of 3D pose estimates in the first three experiments, when the average MAE error of 1.5° (23mm), 4.6° (50mm) and 6.0° (54mm) is reported respectively. However, when inaccurate 3D pose estimates are used (4th experiment) the Isomap (4.8°) outperforms LE (5.0°). In all cases, refined 3D poses obtained with Isomap are systematically more stable, i.e. smaller standard deviation (SD), than LE. When the testing and the training subjects are different (2nd, 3rd and 4th experiments), LLE is much worse and performs only slightly better than PCA if the optimal neighbourhood size is correctly evaluated.

As expected, the accuracy of estimation decreases when subjects differ the most from the one used for training. In

the fourth experiment, where the initial average MAE error of the estimates is 7.0° (79.3mm) with SD 4.5° (41.6mm), after Isomap refinement the error drops to 4.8° (55mm), which is in line with results obtained by projecting MoCap data (S2T1) of the same male subject in the embedded space.

Analysis of the different quantitative measures used to choose 'K', demonstrates the superiority of MI in all conducted experiments. Although it does not always identify the optimal 'K', it outperforms the other metrics and allows the production of results which are not significantly different from those obtained for the best 'K'. An interesting observation is that Isomap and LE seems to be less sensitive to the selection of 'K' than LLE.

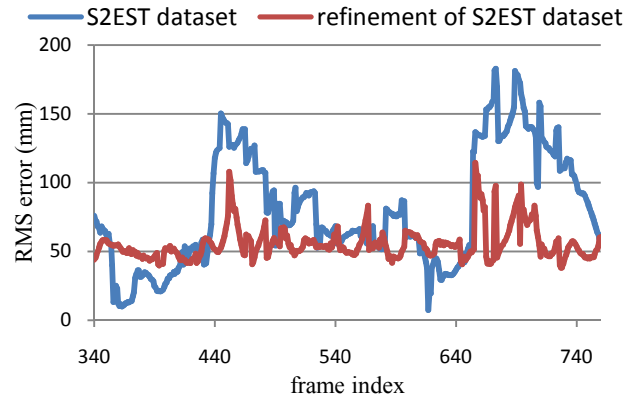


Figure 6: Refinement results of the dataset S2EST for each frame.

It is worth to point out, that all methods enhanced significantly the quality of pose estimates in the fourth experiment which validates the proposed refinement methodology. A more detailed comparison of pose estimates before and after refinement is shown in Figure 6 where ISOMAP was applied using 'K' value predicted by MI. Since we use different subjects for training and testing, the quality of refinement cannot go over a certain threshold which expresses individual differences between walkers. Therefore, for some frames 350-410, 561-568, 605-655 pose estimates are worst after refinement. However, in average accuracy is improved significantly (30%) and our scheme provides much more stability in pose prediction: SD drops from 4.5° (i.e. 41.6mm) to 0.5° (11.8mm). Figure 7 illustrates the effect of our framework by showing refined poses against initial estimates and ground truth.

4.4. Comparison with related research

Since HE datasets are benchmarks for assessing pose recovery methods, we can perform quantitative comparison

of our findings. In [3] an exemplar based approach is used for the 3D pose recovery. Although their initial results show very accurate estimates with errors of 41mm for subject 1 and 39 mm for subject 2, they are obtained when their training set included the evaluated subjects. When subject 1 was removed from training, then its mean error increased to 80mm. This suggests the error they would get for subject 2, when excluded from training, would be well above the 55mm error which we achieve for same subject. Another limitation is the requirement of a huge training database, i.e. almost 20,000 samples, whereas our experiment where conducted using only 1000 frames in our training set.

Our framework has similarities with the work presented in [1; 14], where they report a mean error of 37mm for the walking sequence of subject 2. Although this result is better than ours, it is important to mention that not only subject 2 is used as a part of the training set, but, both the RBFN as well as dimensionality reduction method are tuned. Finally, their approach has additional constraints: it requires dense sampling of training data, which is achieved in practice

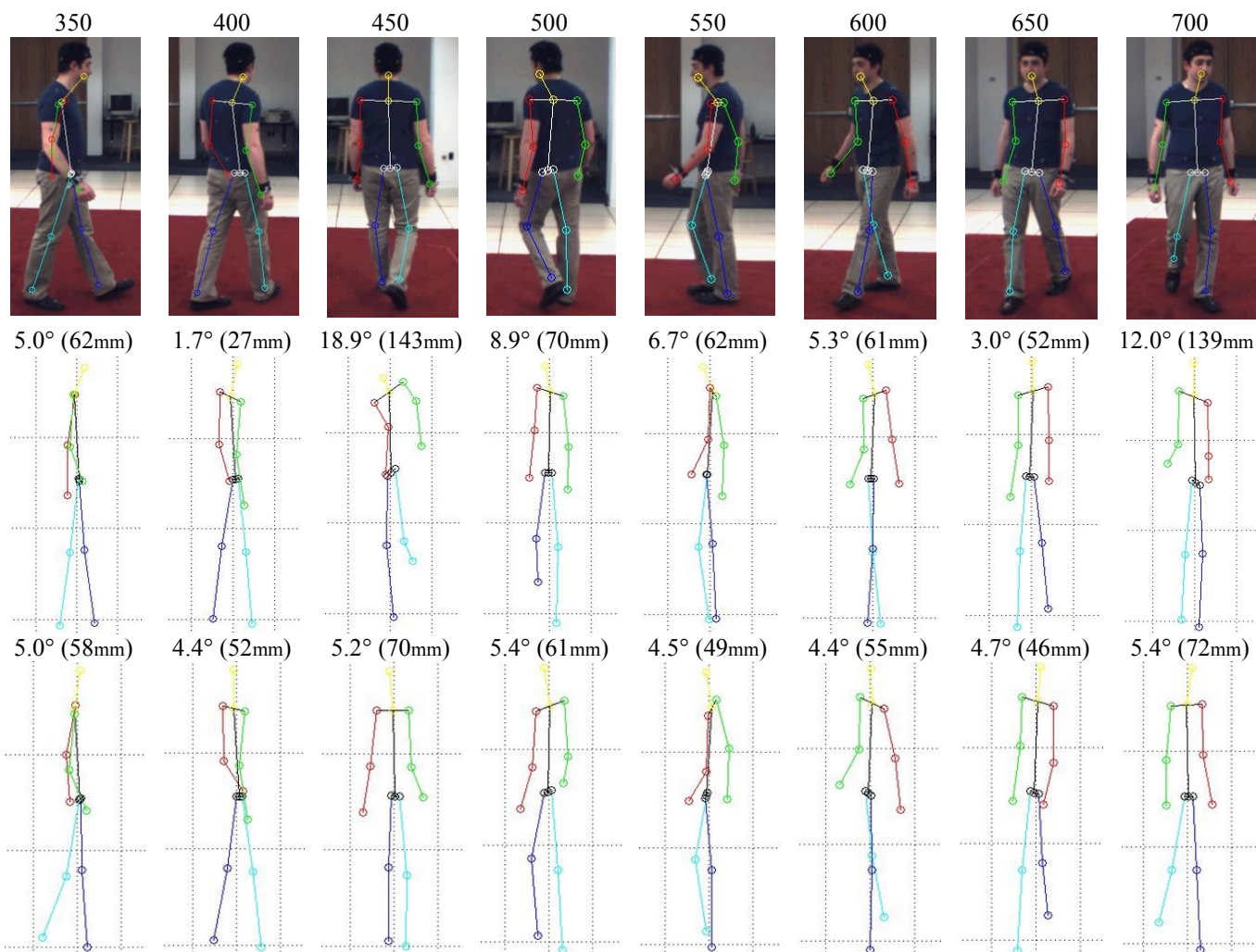


Figure 7: Refinement results; first row: ground truth with frame index; second row: estimated pose with initial MAE and RMS error; third row: refined pose with output MAE and RMS error.

using synthetic training data and it is only suitable for periodic motions.

5. Conclusions and future work

In this paper, the fully automatic configuration of spectral dimensionality reduction methods is proposed and applied to the refinement of 3D pose estimates. The automatic configuration is performed, first, by introducing the new metric MI for estimating neighbourhood size automatically. Secondly, the RBFN, which is exploited to project data between spaces, is learned in a completely unsupervised manner. The effectiveness and stability of the approach was validated in the refinement of 3D pose estimates in a broad range of experiments using a variety of human actors. We validate that our proposed metric MI outperforms previous suggestions such as PM, RV and SR. Also, we identify the Isomap as the best spectral dimensionality reduction method for human motion. In particular, when applied in the experiment with real 3D pose estimates of walking sequence, it allows improving accuracy by 30%. As future research directions, we intend to investigate the further optimization of RBF network structure. Also, we would like to incorporate a dynamical model into the embedded space to constrain the solution of finding nearest manifold representation.

References

- [1] A. Elgammal, C-S. Lee. Nonlinear manifold learning for dynamic shape and dynamic appearance. CVIU. 2007, Vol. 106.
- [2] N.D. Lawrence, C.H. Ek, H.S. Torr. Gaussian process latent variable models for human pose estimation. MLMI. 2007.
- [3] R. Poppe. Evaluating Example-based Pose Estimation: Experiments on the HumanEva sets. CVPR. 2007.
- [4] P. Kuo, A. Thibault, M. Lewandowski, D. Makris, J.-C. Nebel. Exploiting Human Bipedal Motion Constraints for 3D Pose Recovery from a Single Uncalibrated Camera. ISVC. 2009.
- [5] M.W. Lee, I. Cohen. A Model-Based Approach for Estimating Human 3D Poses in Static Images. IEEE TPAMI. 2006.
- [6] F. Caillette, T. Howard. Real-Time Markerless Human Body Tracking with Multi-View 3-D Voxel Reconstruction. In Proc. BMVC. 2004.
- [7] K. Grochow, S. Martin, A. Hertzmann, Z. Popovic. Style-Based Inverse Kinematics. ACM SIGUCCS. 2004.
- [8] C.J. Taylor. Reconstruction of Articulated Objects from Point correspondences in a Single Uncalibrated Image. CVPR. 2000.
- [9] S. Hou, A. Galata, F. Caillette, N. Thacker, P. Bromiley. Real-time Body Tracking Using a Gaussian Process Latent Variable Model. ICCV. 2007.
- [10] A. Kraskov, H. Stogbauer, P. Grassberger. Estimating Mutual Information. E-print, arXiv.org/cond-mat/0305641. 2003.
- [11] T. Poggio, F. Girosi. Networks for Approximation and Learning. Proc. IEEE. 78, 1990, Vol. 9.
- [12] G. Mori, J. Malki. Recovering 3d Human Body Configurations Using Shape Contexts. TPAMI. 2006, vol. 28.
- [13] C. Menier, E. Boyer, B. Raffin. 3D Skeleton-Based Body Pose Recovery. 3DPVT. 2004.
- [14] A. Elgammal, C-S. Lee. Body Pose Tracking From Uncalibrated Camera Using Supervised Manifold Learning. NIPS Workshop on Evaluation of Articulated Human Motion and Pose Estimation. 2006.
- [15] L.K. Saul, S.T. Roweis. Nonlinear dimensionality reduction by locally linear embedding. Science. 2000.
- [16] J.B. Tenenbaum. Global Geometric Framework for Nonlinear dimensionality reduction. Science. 290, 2000.
- [17] M. Belkin, P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Computation. 2003.
- [18] E. Levina, P.J. Bickel. Maximum Likelihood Estimation of Intrinsic Dimension. NIPS. 2005.
- [19] B. Kegl. Intrinsic Dimension Estimation Using Packing. NIPS. 2002.
- [20] J.A. Costa, A.O. Hero. Geodesic Entropic Graphs for Dimension and Entropy Estimation in Manifold Learning. IEEE TSP. 2004.
- [21] O. Kouropteva, O. Okun, M. Pietikainen. Selection of the optimal parameter value for the locally linear embedding algorithm. FSKD. 2002.
- [22] O. Samko, A.D. Marshall, P.L. Rosin. Selection of the optimal parameter value for the Isomap algorithm. PRL. 27, 2006, Vol. 9.
- [23] R. Karbauskaite, O. Kurasova, G. Dzemyda. Selection of the number of neighbours of each data point for the Locally Linear Embedding Algorithm. ITC. 36, 2007, Vol. 4.
- [24] Y. Goldberg, Y. Ritov. Local Procrustes for Manifold Embedding: A Measure of Embedding Quality and Embedding Algorithms. Submitted to J. of Machine Learning. 2008.
- [25] G. Mori, J. Malki. Estimating Human Body Configurations using Shape Context Matching. ECCV. 2002.
- [26] F. Remondino, A. Roditakis. 3D Reconstruction of Human Skeleton from Single Images or Monocular Video Sequences. DAGM03. 2003.
- [27] S. Chen, C.F.N. Cowan, P.M. Grant. Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks. IEEE NN. 1991, Vol. 2, No. 2.
- [28] L. Xu, A. Krzyżak, E. Oja. Rival Penalized Competitive Learning for Clustering Analysis, RBF net, and Curve Detection. IEEE NN. 1993, Vol. 4, No. 4.
- [29] T. Kanungo, D.M. Mount, N. Netanyahu, C. Piatko, R. Silverman, A.Y. Wu. A Local Search Approximation Algorithm for k-Means Clustering. Proc. of the 18th ACM Symp. on Computational Geometry. 2002.
- [30] HumanEVA dataset, Brown University. <http://vision.cs.brown.edu/humaneva>.
- [31] L. Sigal, M. J. Black. HumanEva: Synchronized video and motion capture dataset for evaluation of articulated human motion. Brown University : 2006. Tech. Report CS0608.
- [32] G.A.F. Seber. Multivariate Observations. Wiley. 1984.