Comparison of Shape Regression Methods under Landmark Position Uncertainty

Nora Baka[1,2], Coert Metz[1], Michiel Schaap[1], Boudewijn Lelieveldt[2,4], Wiro Niessen[1,4], and Marleen de Bruijne[1,3]

[1] Erasmus MC, Rotterdam, The Netherlands
n.baka@erasmusmc.nl
[2] Leiden University Medical Center, Leiden, The Netherlands
[3] University of Copenhagen, Denmark
[4] Delft University of Technology, The Netherlands

Abstract. Despite the growing interest in regression based shape estimation, no study has yet systematically compared different regression methods for shape estimation. We aimed to fill this gap by comparing linear regression methods with a special focus on shapes with landmark position uncertainties. We investigate two scenarios: In the first, the uncertainty of the landmark positions was similar in the training and test dataset, whereas in the second the uncertainty of the training and test data were different. Both scenarios were tested on simulated data and on statistical models of the left ventricle estimating the end-systolic shape from end-diastole with landmark uncertainties derived from the segmentation process, and of the femur estimating the 3D shape from one projection with landmark uncertainties derived from the imaging setup. Results show that in the first scenario linear regression methods tend to perform similar. In the second scenario including estimates of the test shape landmark uncertainty in the regression improved results.

1 Introduction

Shape estimation by linear regression has been applied in a great variety of problems, including 3D shape estimation from digitized point cloud [10] and projection images [15], neighboring shape prediction [8,11,14], organ motion prediction based on a few time-points [9,7,1], healthy shape prediction for disease quantification [4], and remaining shape variation prediction from partial field of view [3,2].

The shapes used to construct the regression model are segmentations of images with noise, artifacts, low contrast, etc. Therefore, landmark positions always inhabit a certain amount uncertainty, which may be estimated from the images, imaging setup, segmentation method, etc. Regression methods that are currently applied for shape prediction do not incorporate this additional information.

Studies in chemometrics literature compared the performance of linear regression methods in noisy environments concluding that including data uncertainty may improve prediction in certain situations [12]. These conclusions can not directly be generalized to shape regression, as they apply to single variable outcome

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part II, LNCS 6892, pp. 434–441, 2011.
© Springer-Verlag Berlin Heidelberg 2011

cases with low dimensional input and a large training set. Shape regression on the other hand is characterized by high dimensional input and output variables, and a very small training set.

Therefore, we addressed the following questions: (1) Is there a preferred linear regression method for shape regression? (2) In what context is it beneficial to incorporate landmark position uncertainties in the regression?

We answer these questions by comparing several linear regression methods with and without incorporation of uncertainties. We differentiate two scenarios: In the first, uncertainties of the landmark positions are similar in the training and test dataset, whereas in the second uncertainties of the training and test data are different. We evaluate the performances in simulation experiments providing a controlled setup with a Gaussian shape and uncertainty model and known ground truth, and in a real-world dataset for each scenario. First, the end-systolic left ventricle shape is predicted from end-diastole with landmark uncertainties derived from the segmentation process, and second, the 3D femur shape is estimated from one projection with landmark uncertainties derived from the imaging setup.

2 Shape regression

Shape representation. Each training shape consists of landmark points, which are in correspondence across all shapes. After rigid alignment each shape can be represented by its concatenated landmark coordinates as a high dimensional point. The set of training shapes in this high dimensional space is commonly assumed to be Gaussian distributed with lower intrinsic dimensionality.

Regression techniques. In this paper we describe shape regression where both input $\boldsymbol{x}$ and output $\boldsymbol{y}$ are shapes. We focus on linear regression methods, as in low sample size high dimensionality problems non-linear methods tend to over-fit the training data [6]. Ordinary least squares (OLS) is the best unbiased linear estimator. It minimizes the sum of squares prediction error requiring an over-determined system, and often inhabiting large prediction variance[1]. Due to the small sample size and the lower intrinsic dimensionality shape regression is typically under-determined, and OLS is not directly applicable. We compared biased methods as they both reduce prediction variance, and are applicable for under-determined problems.

Standard biased regressions not including landmark uncertainty are principal component regression (PCR) applying OLS on a sub-space calculated by principal component analysis; partial least squares regression (PLS) optimizing the underlying lower dimensional spaces of input and output to maintain the most covariance between them[2]; and ridge regression (RR) applying a norm constraint on the regression coefficients. See [6] for more detailed descriptions.

[1] With multivariate $\boldsymbol{x}$ and $\boldsymbol{y}$ OLS is termed multiple outcome multiple linear regression.
[2] We used the SIMPLS algorithm for PLS.

Maximum likelihood (ML) PCR includes uncertainties of both test and training data by performing OLS on the input subspace A, that is optimal in the ML sense for a given set of samples with known Gaussian uncertainties. A sample $\boldsymbol{x}$ with uncertainty Ψ is represented as $r = (A^T\Psi^{-1}A)^{-1}A^T\Psi^{-1}\boldsymbol{x}$ in this subspace. We defined a minimum uncertainty for all dimensions to maintain stability. Furthermore to cut computational costs, we used an efficient alternating algorithm with independent noise approximation and no intercept optimization [13]. MLPCR has not been used in shape regression before.

To include uncertainty in the test sample only, we used maximum a-posteriori (MAP) ridge regression, proposed in [2]. With known test uncertainty Ψ one obtains the coefficient matrix $B_{\mathrm{MAPR}} = \Sigma_{YX}(\Sigma_{XX} + \Psi + \lambda I)^{-1}$, where Σ_{XX} and Σ_{YX} are the input covariance and cross-covariance matrices, I is the identity matrix, and λ is a regularization constant.

Suitable values for the percentage of variance retained with PCR, number of modes of PLS and MLPCR, and regularization constant for ridge and MAP ridge regressions have to be assessed from the training set. We used cross-validation on the training set to estimate the regularization.

3 Simulation Study

Framework description. To evaluate shape regression with different noise patterns, we simulated data that resembles real world shapes, i.e. (1) the data forms a low-dimensional Gaussian distribution with variance outside this subspace being Gaussian noise; (2) noise of neighboring dimensions is correlated; (3) the variance of modes drops quickly and then smoothly converges to a constant; (4) sample dimensionality greatly exceeds the training set size; and (5) the input only partially predicts the output.

The data was generated with intercepts $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ as follows

$$\boldsymbol{x}^i = \boldsymbol{x}_0 + \Phi_{in}\boldsymbol{\alpha}^i + \boldsymbol{\delta}^i_{in} \quad , \quad \boldsymbol{y}^i = \boldsymbol{y}_0 + \Phi_{out}(W\boldsymbol{\alpha}^i) + \boldsymbol{\delta}^i_{out} \tag{1}$$

where $\boldsymbol{\alpha}^i$ is the d_{in} dimensional representation of the i-th training input $\boldsymbol{x}^i$, and is sampled per dimension from a Gaussian distribution with decreasing variance. The matrices Φ_{in} and Φ_{out} contain random orthonormal vectors, and W is the random orthogonal linear regression coefficient matrix of size $d_{in} \times d_{out}$. The noise $\boldsymbol{\delta}^i_{in}$ and $\boldsymbol{\delta}^i_{out}$ is sampled from a Gaussian distribution with covariance

$$\Sigma_\delta = \zeta\left[\Sigma_{common} + \Sigma_{sampleSpec}\right] = \zeta\left[\Theta^T\Gamma\Theta + S^TS\right] \quad , \tag{2}$$

where ζ regulates the amount of noise. Noise consists of Σ_{common} producing a shared noise structure, and a sample specific term $\Sigma_{sampleSpec}$. The first is calculated from a random low dimensional space spanned by the columns of Θ with a magnitude defined by the diagonal of Γ, while the second is calculated from a random matrix S consisting of uniform gaussian samples, except of a diagonal band exhibiting larger variance. Such noise models the common uncertainty structure shared by all samples, such as a larger error along the shape surface than perpendicular to it as well as the smoothness of real shapes making noise highly correlated in a neighborhood.

Experiments and results. We conducted experiments with 200 dimensional input and output of intrinsic dimensionality $d_{in} = 5$, $d_{out} = 3$. We chose ζ such that the noise magnitude was 10% of the real data variation. Training set size was 30, unless stated differently. For the first two experiments the exact noise covariance was used as sample uncertainty estimate. We used the root mean squared (RMS) distance of the predicted output and the noise-free test output for evaluation. The reported results are averages of 100 random realizations.

The first experiment simulates applications where training and test shapes are produced with the same segmentation method and imaging protocol. This scenario is most common in real shape regressions. For this simulation training and test input and output is given same amount of noise. Results for different training set sizes shown in Fig. 1 indicate only small differences between regression methods. Results of 2-sample t-tests on the pooled training set sizes are shown in Tab. 1. The same experiment with 50% noise gave similar results.

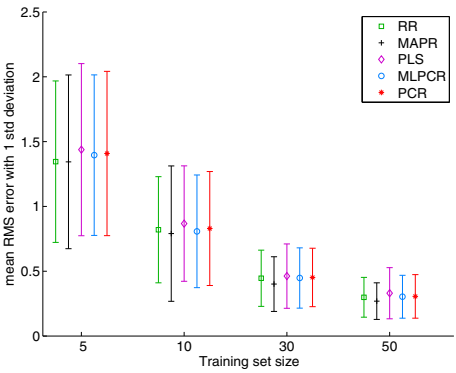

Fig. 1. Error as function of training set size for different regression methods, when noise in training and test data is similar (Simulation experiment 1)

Table 1. Significance table: The upper triangle summarizes results when training and test data are similar (S1 - simulation 1, H - Heart). The lower triangle summarizes experiments with different training and test data (S2 - simulation 2, F - femur 2D-3D). The arrows point to the significantly outperforming method (5% confidence level).

	S1 H	S1 H	S1 H	S1 H
RR	↓ 0	← ←	0 ←	← ←
→ →	MAPR	← ←	← ←	← ←
↑ ↑	↑ ↑	PLS	↓ 0	↓ ↓
0 ↑	↑ ↑	→ ↑	MLPCR	0 ↓
0 ↑	↑ ↑	→ 0	0 →	PCR
S2 F	S2 F	S2 F	S2 F	

The second experiment simulates applications where the training shapes are created from a different modality or with a different method than the test data, e.g. training shapes from CT and test data from X-ray [15]. They exhibit thus different noise structure. We simulated the training set with 10% noise, while test sample noise varied between 5% and 100%. Results are shown in Figure 2 and Table 1.

In the third experiment we assessed the effect of wrongly estimated amount of test uncertainty. We used the setup of the previous experiment with 10% training set noise, and twice as large test noise, and varied the estimated magnitude of the test uncertainty. Note, that the covariance structure of the test uncertainty stayed the same, only the magnitude varied. Results are shown in Figure 3.

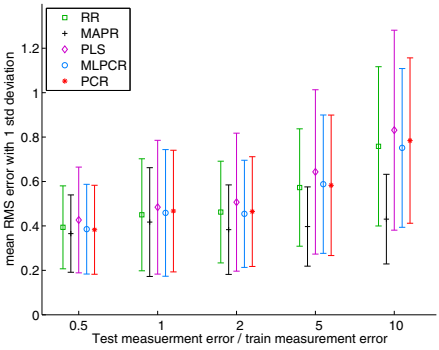

Fig. 2. RMS error for different regression methods as function of the ratio between test and training set noise magnitude (Simulation experiment 2)

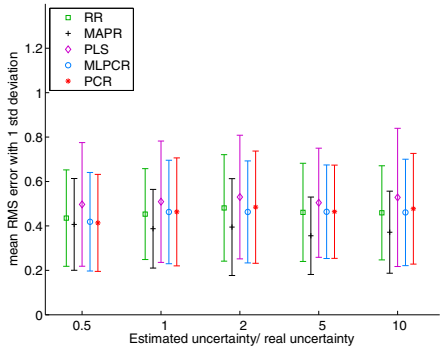

Fig. 3. RMS error for different regression methods as function of the ratio between estimated and actual test noise magnitude (Sim. experiment 3)

4 Real World Datasets

Left ventricle prediction. Estimation of a patient-specific 4D beating heart model is useful for integration of pre-operative 3D CTA in X-ray guided coronary angioplasty procedures [9]. We investigated the accuracy of predicting the end-systolic left ventricle shape from end-diastole as a real world example for the scenario of similar test and train uncertainties.

We extracted 150 end-diastolic and end-systolic left ventricle (LV) shapes from 4D cardiac CTA segmentations, derived by atlas segmentation at end-diastole and subsequent 4D registration as proposed by [9]. The atlases contained 5899 landmark points denoting the LV endo- and epi-cardium. The end-diastolic segmentations were obtained by averaging the landmark points of the eight registered atlases. The spread of these points was used to estimate the landmark uncertainty. Example segmentations with color-coded uncertainty magnitude per landmark are shown in Fig. 4).

The RMS point-to-point (P2P) distance between predicted and segmented end-systolic shape was evaluated in leave-one-out experiments. The regularization parameters in PCR, PLS and RR were optimized via 15-fold cross-validation, and MLPCR was assigned the same number of modes as PCR. Due to the high dimensional input, MAPR was performed in the data subspace, and was optimized per test sample. Results for the different regression methods are shown in Fig. 6 with Table 1 indicating statistical significance.

2D-3D femur reconstruction. Estimation of 3D patient specific bony anatomy from one or more X-Ray images would be beneficial to minimize acquisition costs and radiation dose. Regression based reconstruction after rigid alignment and 2D-3D correspondence generation was proposed by Zheng et al.[15]. We focused on the reconstruction step assuming known projection parameters,

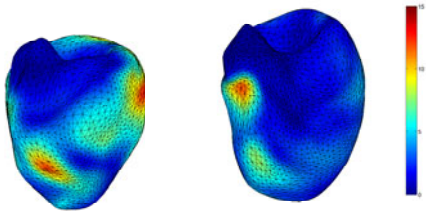

Fig. 4. Two end-diastolic left ventricle shapes with color-coded landmark uncertainty, derived from CTA data by multi-atlas segmentation

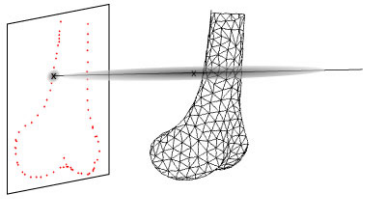

Fig. 5. The 2D-3D uncertainty structure. The 3D position of a projected 2D silhouette point is highly uncertain in the projection direction.

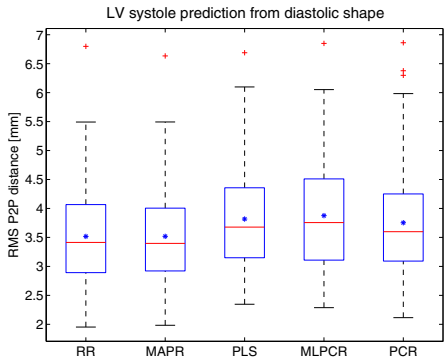

Fig. 6. Box plots showing RMS errors between estimated and reference LV systolic shape. The blue star marks the mean RMS error.

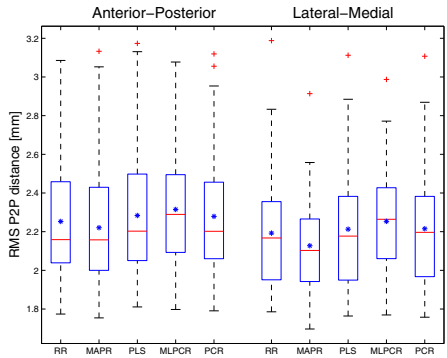

Fig. 7. Box plots showing RMS errors between estimated and reference 3D femur shape for two projection directions

optimal pose, and perfect 2D-3D correspondence. This is an example of the scenario when training and test uncertainty differs.

The regression coefficient matrix was calculated with as input the 3D silhouette landmarks and as output the entire 3D shape. The test input was generated from the 2D projected silhouette landmarks as the 3D point on the projection ray, which is closest to its corresponding landmark of the mean shape. Therefore, the test input was given a large uncertainty along the projection rays, and a small one perpendicular to it, as shown in Fig. 5. The training shapes were rigidly aligned based on the silhouette landmark points only, as proposed in [2]. We performed leave-one-out experiments to estimate the 3D distal femur shape from one projection.

The femur model was created from semi-automatic segmentations of 29 patient CTs and 13 cadavers from both sexes with varying ages. Point correspondence was generated with the GAMES algorithm [5]. The resulting shape model contained 33 modes of variations with 95% retained variance.

We compared the regression methods based on the RMS P2P distance between predicted and test output. Results for the Anterior-Posterior (AP) and the Lateral-Medial (LM) projection directions are shown in Fig. 7 with significance indication in Table 1.

5 Discussion and Conclusions

Despite the growing interest in regression based shape prediction, no study had before systematically compared different regression methods for shape analysis. We aimed to fill this gap by comparing several linear regression methods with a special focus on landmark position uncertainties. We defined two uncertainty scenarios.

In the first scenario test and train uncertainties were similar. Simulation experiment 1 and the LV prediction experiment (Fig. 1, Fig. 6, and Table 1) showed little differences between different linear regression methods. Uncertainty incorporation via MLPCR did not improve results. A similar phenomenon was reported in [12] for MLPCR without intercept optimization. Results might improve without the simplifications of independent noise and no intercept, but the high computational requirements make it currently impractical for shape regression. Including uncertainty of the test sample via MAPR marginally improved predictions in the simulations, but did not improve upon RR on the heart data. While the regularization constant of MAPR depends on the test uncertainty, and has to be re-optimized for every new test sample, the optimization of the standard regression methods is solely based on the training data, can thus be precomputed. We observed that the average cross-validation curve used to chose the shrinkage parameter was smooth for RR, and spiky for PCR and PLS due to the discrete inclusion of directions, and the false correlation magnitudes in case of a small and noisy training set. This might be one reason why RR slightly but consistently outperformed PCR and PLS. Overall, in this scenario we found RR to perform slightly better than to other linear regression methods.

In the second scenario training and test uncertainties differ largely. Simulation experiment 2 and the 2D-3D femur reconstruction experiment (Fig. 2, Fig. 7 and Tab. 1) consistently showed that including knowledge of the test uncertainty in the prediction via MAPR significantly improves results. The practical gain depends on the difference in uncertainty and on how well the assumed normal distribution fits the shape variability and the noise. Real world data might therefore produce a smaller gain with MAPR. Estimation errors of the noise magnitude with the correct covariance structure are, however, tolerable within a large interval (simulation 3, Fig. 3). The performance of RR,PLS,PCR and MLPCR did not change from scenario 1. Therefore, we conclude that in such situations including landmark uncertainty of the test sample may be beneficial.

In our comparison the predicted shape is the end result. Our work could be naturally extended to include the remaining variance of the prediction.

We believe that the presented comparison study gives insights for researchers working on shape prediction in various domains, and helps in the selection of a suitable regression method for future applications.

The Netherlands Organization for Scientific Research (NWO) and SenterNovem, project IGIT4Health is greatly acknowledged for its financial support.

References

1. Ablitt, N.A., Gao, J., Keegan, J., Stegger, L., Firmin, D.N., Yang, G.Z.: Predictive cardiac motion modeling and correction with partial least squares regression. IEEE Trans. Med. Imaging 23(10), 1315–1324 (2004)
2. Baka, N., de Bruijne, M., Niessen, W., Reiber, J.H.C., Lelieveldt, B.: Confidence of model based shape reconstruction from sparse data. In: IEEE ISBI (2010)
3. Blanc, R., Syrkina, E., Székely, G.: Estimating the confidence of statistical model based shape prediction. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 602–613. Springer, Heidelberg (2009)
4. de Bruijne, M., Lund, M.T., Tankó, L.B., Pettersen, P.C., Nielsen, M.: Quantitative vertebral morphometry using neighbor-conditional shape models. Med. Image Anal. 11(5), 503–512 (2007)
5. Ferrarini, L., Olofsen, H., Palm, W., van Buchem, M., Reiber, J., Admiraal-Behloul, F.: Games: growing and adaptive meshes for fully automatic shape modeling and analysis. Med. Image Anal. 11, 302–314 (2007)
6. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning (Data mining, Inference, and Prediction). Springer, Heidelberg (2009)
7. Klinder, T., Lorenz, C., Ostermann, J.: Prediction framework for statistical respiratory motion modeling. In: Jiang, T., Navab, N., Pluim, J., Viergever, M. A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 327–334. Springer, Heidelberg (2010)
8. Liu, T., Shen, D., Davatzikos, C.: Predictive Modeling of Anatomic Structures Using Canonical Correlation Analysis. In: IEEE ISBI, pp. 1279–1282 (2004)
9. Metz, C., Baka, N., Kirisli, H., Schaap, M., van Walsum, T., Klein, S., Neefjes, L., Mollet, N., Lelieveldt, B., de Bruijne, M., Niessen, W.: Conditional shape models for cardiac motion estimation. In: Jiang, T., Navab, N., Pluim, J., Viergever, M. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 452–459. Springer, Heidelberg (2010)
10. Rajamani, K.T., Styner, M.A., Talib, H., Zheng, G., Nolte, L.P., Ballester, M.A.: Statistical deformable bone models for robust 3d surface extrapolation from sparse data. Medical Image Analysis 11(2), 99–109 (2007)
11. Rao, A., Aljabar, P., Rueckert, D.: Hierarchical statistical shape analysis and prediction of sub-cortical brain structures. Med. Image Anal. 12(1), 55–68 (2008)
12. Reis, M.S., Saraiva, P.M.: A comparative study of linear regression methods in noisy environments. J. Chemometrics 18, 526–536 (2004)
13. Wentzell, P.D., Andrews, D.T., Hamilton, D.C., Klaas, F., Kowalski, B.R.: Maximum likelihood principal component analysis. J. Chemometrics 11, 339–366 (1997)
14. Yang, Y.M., Rueckert, D., Bull, A.M.J.: Predicting the Shapes of Bones at a Joint: Application to the Shoulder. Computer Methods in Biomechanics and Biomedical Engineering 11(1), 19–30 (2008)
15. Zheng, G., Gollmer, S., Schumann, S., Dong, X., Feilkas, T., Ballester, M.A.G.: A 2d/3d correspondence building method for reconstruction of a patient-specific 3d bone surface model using point distribution models and calibrated x-ray images. Med. Image Anal. 13(6), 883–899 (2008)