MODEL-ORDER SELECTION IN STATISTICAL SHAPE MODELS

Alma Eguizabal, Peter J. Schreier

Signal and System Theory Group
University of Paderborn, Germany
{alma.eguizabal, peter.schreier}@sst.upb.de

David Ramírez

Dept. of Signal Theory and Communications
Universidad Carlos III de Madrid, Spain
david.ramirez@uc3m.es

ABSTRACT

Statistical shape models enhance machine learning algorithms providing prior information about deformation. A Point Distribution Model (PDM) is a popular landmark-based statistical shape model for segmentation. It requires choosing a model order, which determines how much of the variation seen in the training data is accounted for by the PDM. A good choice of the model order depends on the number of training samples and the noise level in the training data set. Yet the most common approach for choosing the model order simply keeps a predetermined percentage of the total shape variation. In this paper, we present a technique for choosing the model order based on information-theoretic criteria, and we show empirical evidence that the model order chosen by this technique provides a good trade-off between over- and underfitting.

Index Terms— Model-order selection, information-theoretic criteria, statistical shape model.

1. INTRODUCTION

Statistical shape models provide prior information about the deformation of an object [1]. A Point Distribution Model (PDM) [2] contains statistical information of a collection of shape landmarks and its variability, represented by an affine space of eigenvectors obtained by Principal Component Analysis (PCA). Popular segmentation techniques that make use of PDMs are Active Shapes Models [3] and Constrained Local Models [4]. We refer to the number of principal components kept in the PDM as its model order. If the order is too large, the model may not be specific enough (overfitting); if the order is too small, new observations of the same shape may not

be accurately represented with the model (underfitting). A good model order provides the right trade-off between overfitting and underfitting.

The most common way of choosing the model order of a PDM is to keep the eigenvectors that account for a given percentage of variance (typically 90-98% [4]). Many landmark-based shapes, such as anatomical shapes in medical image analysis, are high-dimensional, and often only few observations are available. These observations may also contain noise artifacts. The heuristic approach of choosing the model order based on a kept fraction of total variance may therefore be suboptimal. The best model order varies significantly depending on the number of samples and noise level of the training data set. Our motivation in this paper is to design a model-order selection rule that has a theoretical justification and leads to a statistical shape model with good representation ability. We consider therefore information theory, which has successfully been used before to enhance registration and detection algorithms [5][6], as well as to place landmarks automatically in statistical shape models [7].

This model-order selection problem has been addressed before in statistical shape model design. In [8] the authors suggested a t-test of bootstrap stability of the PCA modes of the PDM, and they validated the strategy on simulated anatomical shapes with white noise. The authors in [9] proposed a strategy based on an information-theoretic criterion for small sample support in a more generic array-processing context, also assuming white noise. These techniques may fail if the noise is not white. In [10], also in an array-processing context, the authors considered nonuniform noise.

In PDM design, there is no obvious model for the noise. Therefore, techniques that consider a specific noise structure may not work well. In order to address this, we propose a new strategy, based on information-theoretic criteria, that assumes a more generic colored-noise model. Our strategy is specifically designed to determine the model order in a PDM, although it may also be applied to other model-order selection problems with colored noise. We interpret the PDM as a multivariate regression, where the model order is determined considering the statistical properties of the regression residuals. We also perform a comparative study, with simulated

and real shapes, where we prove the good performance of our strategy, as well as the importance of an accurate model order in PDMs.

The remaining of the paper is organized as follows: Section 2 explains in detail the problem formulation, and how the PDM is interpreted as a regression problem. Section 3 presents the proposed solution based on information theory. Section 4 shows a comparative and validation study with simulated shapes where the model order is known, and shapes from real databases where the model order is unknown. We summarize the conclusions in Section 5.

2. PROBLEM FORMULATION

A PDM of an object of interest models the variability of its shape [2]. Let the random vector $\mathbf{x} \in \mathbb{R}^N$ model the x- and y-coordinates of $\frac{N}{2}$ landmarks of a shape after a Procrustes aligment [11]. With the mean $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}_{\mathbf{x}}$, we define the zero-mean shape vector as $\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}$.

The N variables in $\mathbf{y}$ are highly correlated, and most of the information on shape variability is contained in a lower-dimensional subspace. Let us define the covariance matrix of the shape vector, $\mathbf{R}_{yy} = \mathbb{E}[\mathbf{y}\mathbf{y}^T]$, and consider its eigenvalue decomposition, i.e., $\mathbf{R}_{yy} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T$. We perform dimensionality reduction for $\mathbf{y}$, assuming the following linear model:

$$\mathbf{y} = \mathbf{P}_t\mathbf{b}_t + \boldsymbol{\epsilon}, \qquad (1)$$

where $\mathbf{b}_t \in \mathbb{R}^t$ is a vector of $t < N$ parameters, and $\mathbf{P}_t = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_t] \in \mathbb{R}^{N \times t}$ are the t eigenvectors of $\mathbf{R}_{yy}$ corresponding to the t largest eigenvalues. In order to preserve shape plausibility, $\mathbf{b}_t$ is restricted to a set $\mathbb{B}(\boldsymbol{\lambda}_t)$, where $\boldsymbol{\lambda}_t$ is a t-dimensional vector containing the t largest eigenvalues of $\mathbf{R}_{yy}$. The vector $\boldsymbol{\epsilon}$ accounts for the variability that we do not represent with the model. We consider $\boldsymbol{\epsilon}$ to be a zero-mean Gaussian noise vector, with arbitrary covariance matrix $\boldsymbol{\Sigma}$, i.e., colored noise. The noise may be due to imprecise training landmarks, presence of non-representative variability in the training set, insufficient samples, or quantization errors. Thus, the noise at different landmarks may be correlated and may have different variances. Furthermore, the prior Procrustes alignment typically introduces color in the noise.

The problem is to determine the value of the model order t. Following the definition of best shape model provided in [7], we consider an information-theoretic approach to selecting t, which provides a trade-off between compactness (size of t), specificity (overfitting) and generalization ability (underfitting).

3. PROPOSED SOLUTION

Let us assume we obtain M observations of the vector $\mathbf{x}$, that is, M shape training samples $\mathbf{x}^{(m)}$, $m = 1, \ldots, M$, after a Procrustes alignment. Our solution divides the problem into two steps. First, we obtain estimates for the model parameters $\mathbf{P}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\mu}_{\mathbf{x}}$. Then, we estimate the model order t. For this, we split the observed data into two subsets of sizes M_1 and M_2, denoted by $\mathbf{X}_1 = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(M_1)}] \in \mathbb{R}^{N \times M_1}$ and $\mathbf{X}_2 = [\mathbf{x}^{(M_1+1)}, \ldots, \mathbf{x}^{(M_1+M_2)}] \in \mathbb{R}^{N \times M_2}$. In our implementation, we choose $M_1 = M_2$.

Within the first set $\mathbf{X}_1$, we compute the sample mean $\hat{\boldsymbol{\mu}}_{\mathbf{x}} = \frac{1}{M_1} \sum_{m=1}^{M_1} \mathbf{x}^{(m)}$ as well as the sample covariance matrix $\hat{\mathbf{R}}_{yy} = \frac{1}{M_1}(\mathbf{X}_1\mathbf{X}_1^T - \hat{\boldsymbol{\mu}}_{\mathbf{x}}\hat{\boldsymbol{\mu}}_{\mathbf{x}}^T)$ and its eigenvalue decomposition $\hat{\mathbf{R}}_{yy} = \hat{\mathbf{P}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{P}}^T$.

Let us define the matrix $\mathbf{Y}$ that contains the entries of $\mathbf{X}_2$ with mean removed, and consider the following multivariate linear regression:

$$\mathbf{Y} = \hat{\mathbf{P}}_t\mathbf{B}_t + \mathbf{E}, \qquad (2)$$

where the columns in $\mathbf{B}_t \in \mathbb{R}^{t \times M_2}$ are parameter vectors $\mathbf{b}_t$ as defined in (1) corresponding to each of the M_2 vectors in set $\mathbf{X}_2$; we model the regression noise in matrix $\mathbf{E} = [\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(M_2)}]$, whose columns are considered independently and identically distributed (i.i.d.) observations of the error vector in (1). We propose an information-theoretic formulation for model selection in multivariate linear regressions, similarly as described in [12]. The model order t^* is chosen as

$$t^* = \arg\min_t (\underbrace{-\log p(\mathbf{Y}|\hat{\mathbf{B}}_t, \hat{\boldsymbol{\Sigma}}_t)}_{\text{likelihood term}} + \underbrace{\eta(t)}_{\text{penalty}}), \qquad (3)$$

where $\hat{\mathbf{B}}_t$ and $\hat{\boldsymbol{\Sigma}}_t$ are the Maximum Likelihood (ML) estimates of the model parameters for model order t, and $\eta(t)$ is a penalty term that depends on the selected information criterion (Akaike, Bayesian, etc.) [13]. We model $\mathbf{B}_t$ as an unknown deterministic parameter. Thus, the only random quantity in the regression is the residual noise matrix $\mathbf{E} = \mathbf{Y} - \hat{\mathbf{P}}_t\mathbf{B}_t$. Consequently, the log-likelihood expression in (3) can be written as [13]:

$$\log p(\mathbf{Y}|\hat{\mathbf{B}}_t, \hat{\boldsymbol{\Sigma}}_t) = -\frac{M_2}{2} \log |\hat{\boldsymbol{\Sigma}}_t|$$
$$-\frac{1}{2}\text{Tr}\{(\mathbf{Y} - \hat{\mathbf{P}}_t\hat{\mathbf{B}}_t)^T\hat{\boldsymbol{\Sigma}}_t^{-1}(\mathbf{Y} - \hat{\mathbf{P}}_t\hat{\mathbf{B}}_t)\} + \text{constant}. \quad (4)$$

3.1. ML estimation of the regression parameters

According to the PDM definition, the columns of $\mathbf{B}_t$, denoted by $\mathbf{b}_t^{(m)}$ for $m = 1, \ldots, M_2$, are contained in a set $\mathbb{B}(\hat{\boldsymbol{\lambda}}_t)$. Thus, the ML estimate of $\mathbf{B}_t$ is also constrained. We define $\mathbb{B}(\hat{\boldsymbol{\lambda}}_t) = \{\mathbf{b}_t \in \mathbb{R}^{t \times 1} : |b_i| \leq \sqrt{\hat{\lambda}_i}, \forall i = 1 \ldots t\}$, where $\mathbf{b}_t = [b_1, \ldots, b_t]^T$, and $\hat{\boldsymbol{\lambda}}_t = [\hat{\lambda}_1, \ldots, \hat{\lambda}_t]^T$ contains the t largest eigenvalues of $\hat{\mathbf{R}}_{yy}$. Then, the ML estimation of $\mathbf{B}_t$ is equivalent to the following regularized least-squares minimization [12]:

$$\hat{\mathbf{B}}_t = \arg\min_{\mathbf{B}_t \in \mathbb{B}(\hat{\boldsymbol{\lambda}}_t)} \text{Tr}\{(\mathbf{Y} - \hat{\mathbf{P}}_t\mathbf{B}_t)^T\boldsymbol{\Sigma}_t^{-1}(\mathbf{Y} - \hat{\mathbf{P}}_t\mathbf{B}_t)\}, \quad (5)$$

where $\mathbf{B}_t \in \mathbb{B}(\hat{\boldsymbol{\lambda}}_t)$ is applied column-wise. We calculate the solution to (5) following the lines of [2]. That is, we first obtain the unconstrained solution to (5), i.e., $\hat{\mathbf{B}}_t^u = (\mathbf{P}_t^T \boldsymbol{\Sigma}_t^{-1} \mathbf{P}_t)^{-1} (\mathbf{P}_t^T \boldsymbol{\Sigma}_t^{-1}) \mathbf{Y}$. Then, we scale it such that the constraints are fullfilled.

The ML estimate of the covariance matrix of the residual noise $\boldsymbol{\Sigma}_t$ is, as long as $N < M_2$, the sample covariance matrix:

$$\hat{\boldsymbol{\Sigma}}_t = \frac{1}{M_2}(\mathbf{Y} - \hat{\mathbf{P}}_t\mathbf{B}_t)(\mathbf{Y} - \hat{\mathbf{P}}_t\mathbf{B}_t)^T. \qquad (6)$$

We observe that the ML estimates $\hat{\mathbf{B}}_t$ and $\hat{\boldsymbol{\Sigma}}_t$ are mutually dependent, which prohibits finding a closed-form solution. We propose an alternating optimization algorithm to find a locally optimum solution. We set $\hat{\boldsymbol{\Sigma}}_t = \mathbf{I}$ as the initial point. Then, for each iteration of the alternating optimization, we solve (5) to calculate the ML estimate of $\hat{\mathbf{B}}_t$ and then, re-estimate the covariance matrix as in (6). We repeat this procedure until convergence.

3.2. Choosing the model order

The shape data is often high-dimensional but with small number of samples. Under these circumstances, the estimate of the matrix $\boldsymbol{\Sigma}_t$ may be ill-conditioned. In order to deal with this, we reduce the number of parameters to be estimated by assuming $\boldsymbol{\Sigma}_t$ to be a diagonal matrix, with $\boldsymbol{\sigma}^2 = [\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2]$ on the diagonal. Consequently, its ML estimate is $\hat{\boldsymbol{\Sigma}}_t = \mathrm{diag}(\frac{1}{M_2}(\mathbf{Y} - \hat{\mathbf{P}}_t\mathbf{B}_t)(\mathbf{Y} - \hat{\mathbf{P}}_t\mathbf{B}_t)^T)$. Following the lines of [12], we choose the Akaike information criterion (AIC), so the penalty term in (3) is $\eta(t) = M_2 t + N$, which corresponds to the degrees of freedom in (4). Notice that AIC has been used as well for similar problems (for example in [9]), and that the assumptions made to derive it are quite weak [13]. Finally, the model-order estimate t^* is obtained by minimizing the terms in (3) that depend on t:

$$t^* = \arg\min_t \left[M_2\left(\sum_{i=1}^{N} \log(\hat{\sigma}_i^2) + 2t\right) \right.$$

$$\left. + \sum_{i=1}^{N}\sum_{m=1}^{M_2} \frac{\hat{\epsilon}_i^{2(m)}}{\hat{\sigma}_i^2} \right], \quad (7)$$

where $\hat{\epsilon}_i^{(m)}$ is the ith element of vector $\hat{\boldsymbol{\epsilon}}^{(m)} = \mathbf{y}^{(m)} - \hat{\mathbf{P}}_t\hat{\mathbf{b}}_t^{(m)}$; and $\hat{\mathbf{b}}_t$ and $\hat{\sigma}_i^2$ are the estimated values after convergence of the alternating optimization.

4. RESULTS AND DISCUSSION

In this section we validate our strategy with numerical results. In a PDM trained from real data sets, the true model order of the shape model is not known. Hence, we have also used simulated shape data following the model in Eq. (1), where the model order t is set beforehand and thus known.
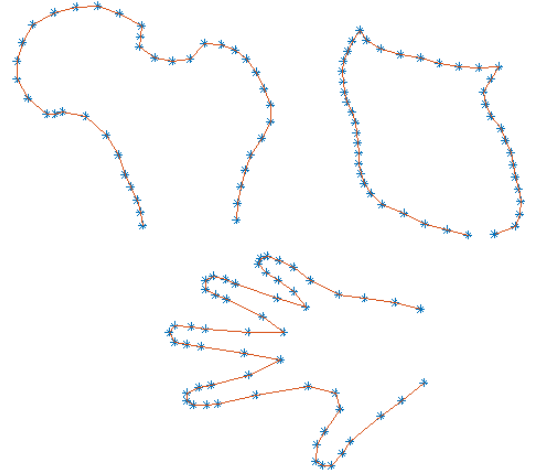

Fig. 1: Average shapes of femur, lung, and hand. The landmarks are shown as blue stars.

4.1. Data description

We simulate realistic synthetic shapes following the model in (1), similarly to the simulated data in [8]. We use the eigenvectors $\hat{\mathbf{P}}_t$ obtained from available real shape data sets and choose values for $\mathbf{b}_t$ that are consistent with the sample data. Then we add white Gaussian noise $\boldsymbol{\epsilon}$ with different noise levels β, which we define as the ratio between the smallest kept signal eigenvalue and the noise variance. We choose these noise levels β such that the produced shapes still look realistic. Then we randomly rotate, scale, and translate these synthetic shapes and use Procrustes to re-align them. Procrustes alignment typically colors the noise, so it may not longer be white when the model order has to be selected.

We have used the following databases of real 2D shapes:

- The top part of the femur bone as seen in fluoroscopic X-rays [14] (168 samples with 40 landmarks).

- Lung outlines from chest X-rays [15] (246 samples with 44 landmarks).

- Hand outlines from photographs [16] (38 samples with 20 landmarks).

We show the average shape of each data set in Fig. 1.

These three data sets are similar: they belong to human anatomy shapes, each sample contains a fixed number of landmarks in correspondence, and the noise is unknown. In each data set, the shapes are composed by a few landmarks that are anatomical and manually labeled, and the rest of the landmarks are equally distributed between these. We refer to the papers [14],[15] and [16] for further details about these data sets.

4.2. Evaluation of the results

We compare our proposed strategy with four others:

1. A variance threshold of 95%, as described in [2] and validated in [4].

2. An information-theoretic technique that considers white noise [9].

3. An information-theoretic criterion considering non-uniform noise [10].

4. A boostrap t-test designed for PDMs [8] that considers white noise.

4.2.1. Simulated data

The evaluation of simulated data is straightforward since there is a known ground truth for the model order t.

In Fig. 2 we show the model order obtained from 1000 Monte Carlo simulations. We see that the performance of the 95%-approach [4] (triangle, yellow lines) depends considerably on the level of noise: there is a tendency to overestimate if β is moderate (5 dB) and to underestimate if β is high (20 dB). The white-noise strategy in [9] (circles, red lines) tends to overestimate, especially when the number of samples increases. We believe this is due to color in the noise, introduced in the simulation by the Procrustes alignment. The t-test stategy [8] (stars, green lines) does not perform well and leads to results with high variance. The reason for this may be that this approach assumes white noise to evaluate stability. The nonuniform noise strategy [10] (diamond, purple lines) tends to underestimate if β is moderate (5 dB), and it provides an incorrect estimation if the number of samples is very small. Our proposed strategy (square, blue lines) outperforms the competitors: it provides the best model-order estimate in general, it needs fewer samples to find the correct estimate, it is not highly dependent on the noise level, and it is consistent with increasing number of samples.

4.2.2. Real data

There is no known ground truth for model order t in a PDM that is trained with real shape data sets. Nevertheless, we may still evaluate how plausible the model-order estimate is when the number of available samples changes and compare this with the behavior in artificial data. Additionally, we illustrate the importance of the model order in PDMs with a numerical experiment that shows the impact of the selection of t when a PDM of order t is used to deal with partial occlusions in shapes.

In Fig. 3 we show the estimated model order for different number of samples on the three data sets. There are similarities with Fig. 2: "95%" [4], t-test [8], and "nonuniform" [10] provide small model orders, while "white-noise" [9] provides
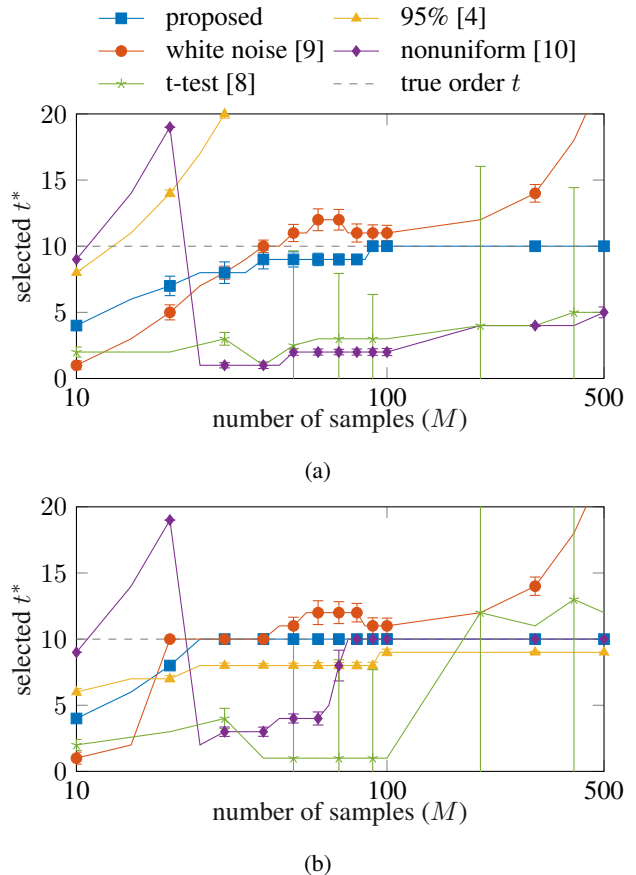

(a)

(b)

Fig. 2: Average model order over 1000 Monte Carlo simulations, using simulated femur shapes of 40 landmarks, with different number of samples M. The vertical bars indicate variance. True model order is $t = 10$. We consider two different signal-to-noise levels: (a) $\beta = 5$ dB. (b) $\beta = 20$ dB.

large (probably too large) model orders. Our strategy seems to provide a consistent model order that starts converging with fewer samples.

As an illustrative example of the importance of the model order in statistical shape models, we performed an experiment that shows how well the PDM with the selected model order can deal with partial occlusions. Considering an obtained PDM as the prior information about shape deformation, we perform an estimate of an occluded (or missing) landmark in a new observed shape. The test consists in the following: inside a leave-one-out test, this is, for all $m = 1, \ldots, M$ available samples in a data set, within the mth "left-out" sample we delete one landmark from that shape. Then, we estimate it from the remaining landmarks using a linear minimum mean-squared error (LMMSE) estimator. The $M - 1$ "not left-out" samples are used to design the PDM, in which we evaluate all possible orders t. Let $\mathbf{y}_i$ denote the missing landmark (which consists of its x- and y-coordinates) and $\mathbf{y}_a$ the remaining available landmarks. Let $\hat{\mathbf{R}}_{ia}$ denote the sample
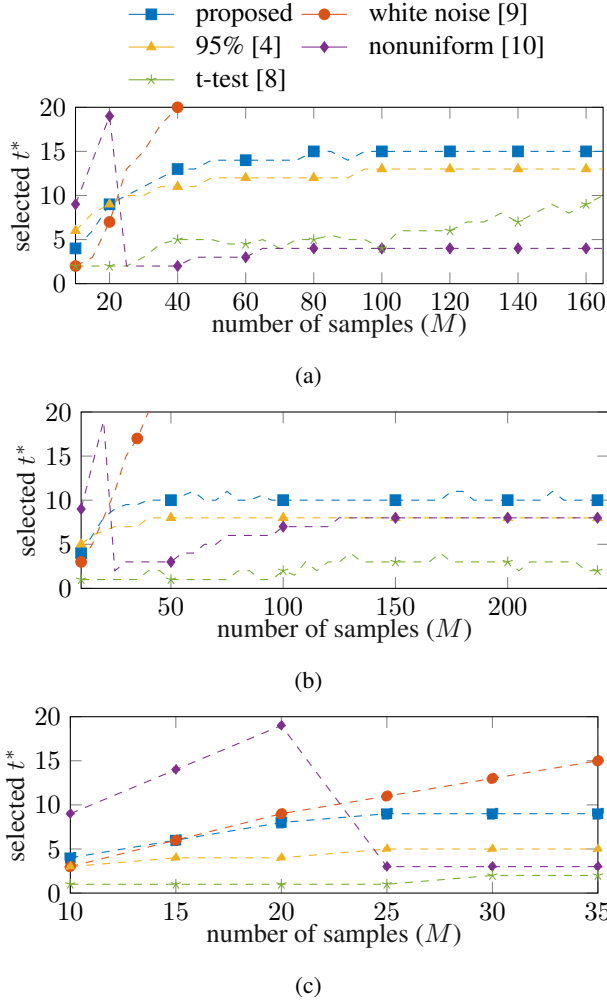
(a)

(b)

(c)

Fig. 3: Selected model order for three different real shape data sets: (a) femur, (b) lung, and (c) hand. Notice that the different shapes have different number of total samples M.

Fig. 4: e_{LMMSE} for all possible order t. The evaluated model orders are obtained using all available samples to train the PDMs. The order selected by "white noise" (orange circle) for the lung data set is not depicted because is out of plot limits.

cross-covariance matrix between the missing landmark and the remaining available landmarks, and $\hat{\mathbf{R}}_{aa}$ the covariance matrix of the available landmarks. These matrices are calculated from the available PDM of order t. The LMMSE estimator of the missing landmark from the remaining landmarks is then $\hat{\mathbf{y}}_i = \hat{\mathbf{R}}_{ia}\hat{\mathbf{R}}_{aa}^{-1}\mathbf{y}_a$. In our experiment, we successively estimate one landmark i from the others, repeating this for all $i = 1, ..., N/2$ landmarks. We average the error over all landmarks and over all available samples and obtain

$$e_{\text{LMMSE}}(t) = \frac{1}{M}\frac{1}{N/2}\sum_{m=1}^{M}\sum_{i=1}^{N/2}||\hat{\mathbf{y}}_i^{(m)} - \mathbf{y}_i^{(m)}||^2, \quad (8)$$

which is evaluated for all possible model orders, i.e., $t = 1, \ldots, \min(N, M)$. Figure 4 shows this metric for the three data sets as a function of considered model order t. In Fig. 4, we observe that the evaluated empirical error decreases until
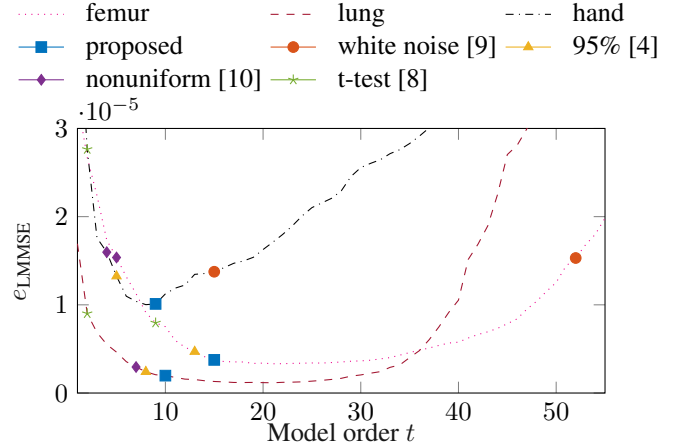
it reaches a minimum, which is different for each data set. A shape model with too small an order may suffer from underfitting (thus, the error decreases if we add more complexity), and too large an order may lead to overfitting (and therefore, the error decreases if we reduce the complexity). We conclude that choosing the right model order is critical in order to minimize the LMMSE. We see that the model order determined by our technique (blue squares) leads to the smallest LMMSE among all competing techniques. We also notice that there is a relatively large interval of model orders that lead to similar LMMSEs. The principle of parsimony dictates that in such a case a smaller order is to be preferred. Our technique observes this principle.

5. CONCLUSION

Statistical shape models provide important information to machine learning algorithms about object deformation. The order of these models is typically obtained heuristically. We have proposed a model-order selection strategy that is based on information-theoretic criteria and thus has a theoretical justification. We have validated the selection performance of our technique on simulated shape data, and it outperformed other model-order selection strategies under different conditions of sample support and noise level. We have also evaluated the technique on shapes from real data sets, showing results similar to the evaluation with artificial data. Additionally, we have performed an empirical test to illustrate the impact of the model order of shape models, and how the choice of order provided by our technique results on a model with better performance.

7. REFERENCES

[1] M. Lüthi, C. Jud, and T. Vetter, "A unified approach to shape model fitting and non-rigid registration," in *Machine Learning in Medical Imaging*, Cham, 2013, pp. 66–73, Springer International Publishing.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models–their training and application," *Computer Vision and Image Understanding*, vol. 61, pp. 38 – 59, January 1995.

[3] C. Santiago, J. C. Nascimento, and J. S. Marques, "2D segmentation using a robust Active Shape Model with the EM algorithm," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2592–2601, Aug 2015.

[4] C. Lindner, S. Thiagarajah, J. M. Wilkinson, The arcO-GEN Consortium, G. A. Wallis, and T. F. Cootes, "Fully automatic segmentation of the proximal femur using random forest regression voting," *IEEE Transactions on Medical Imaging*, vol. 32, pp. 1462–1472, August 2013.

[5] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, Sep 1997.

[6] D. S. Zois and M. Raginsky, "Active object detection on graphs via locally informative trees," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, Sept 2016, pp. 1–6.

[7] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "A minimum description length approach to statistical shape modeling," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 525–537, May 2002.

[8] L. Mei, M. Figl, D. Rueckert, A. Darzi, and P. J. Edwards, "Statistical shape modelling: How many modes should be retained?," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2008.

[9] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Transactions on Signal Processing*, vol. 56, pp. 2625–2638, July 2008.

[10] S. Aouada, A. M. Zoubir, and C. M. Samson See, "Source detection in the presence of nonuniform noise," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2004, vol. 2, pp. 165–168.

[11] F. L. Bookstein, "Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape," *Medical Image Analysis*, vol. 1, pp. 225 – 243, 1997.

[12] E. J. Bedrick and C. L. Tsai, "Model selection for multivariate regression in small samples," *Biometrics*, vol. 50, pp. 226–231, 1994.

[13] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, pp. 36–47, July 2004.

[14] A. Eguizabal and P. J. Schreier, "A weighting strategy for Active Shape Models," in *Proc. of the IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 3610–3614.

[15] S. Junji, I. Shigehiko, K. Junpe, M. Tsuneo, K. Takeshi, K. Ken-ichi, M. Mitate, F. Hiroshi, K. Yoshie, and D. Kunio, "Development of a digital image database for chest radiographs with and without a lung nodule," *American Journal of Roentgenology*, vol. 174, pp. 71–74, 2000.

[16] M. B. Stegmann and D. D. Gomez, "A brief introduction to statistical shape analysis," Tech. Rep., University of Denmark (DTU), 2002.