Correlating Shape and Functional Properties Using Decomposition Approaches

Daniel Dornbusch and **Robert Haschke**
CoR-Lab, Bielefeld University
Universitaetsstr. 25
D-33615 Bielefeld, Germany
ddornbus@cor-lab.uni-bielefeld.de

Stefan Menzel and **Heiko Wersing**
Honda Research Institute Europe GmbH
Carl-Legien-Str. 30
D-63073 Offenbach/Main, Germany

Abstract

In this paper, we propose the application of standard decomposition approaches to find local correlations in multimodal data. In a test scenario, we apply these methods to correlate the local shape of turbine blades with their associated aerodynamic flow fields. We compare several decomposition algorithms, i.e., k-Means, Principal Component Analysis, Non-negative Matrix Factorization and Uni-orthogonal Non-negative Matrix Factorization, with regards to their efficiency at finding local, relevant correlations and their ability to predict one modality from another.

1. Introduction

The relationship between local and global structural properties is a key issue in many complex problem domains, e.g., visual perception, motor planning, and data mining. Methods that are capable of identifying local decompositions of large problems, with possibly several free parameters, into more manageable local elements, are important in order to achieve structural robustness and generalization. Unsupervised decomposition algorithms have been investigated extensively in recent years in the field of object and pattern recognition. In many cases, they effectively aid the completion of tasks by amplifying relevant information (20), filtering out noise (20), demixing and preprocessing signals (3), analyzing data (14), or selecting proper features (10).

Decomposition approaches are capable of identifying statistically relevant correlations in high-dimensional data sets. Hence, we can detect meaningful correlations between *different types* of data, if we apply the algorithms to a unified data set comprising all modalities. As an example we consider turbine rotor blades with various shapes and their associated aerodynamic flow fields. Given a large set of shapes, each one represented as a 2D binary image, decomposition approaches can easily extract basic shape constituents reoccurring in many of the images. Similarly, we can identify common pressure profiles representing the aerodynamic flow field in the vicinity of a blade. However, the obtained parts-based representations of shapes and pressure profiles are typically not meaningful in the context of aerodynamic optimization. This is due to the fact that the extracted basis components of both modalities are not correlated. Nonetheless, applying a decomposition algorithm to a conjoint data set comprising both modalities simultaneously, allows for the identification of inter-modal, semantically meaningful correlations. Also, the presence of another modality can suppress the extraction of functionally irrelevant, but frequently occurring shape constituents, if the associated pressure profiles are uncorrelated with the occurrence of this shape primitive. Hence, the simultaneous use of multimodal data in decomposition approaches can: (i) improve the interpretability of the extracted basis components of each single modality, and (ii) extract functionally relevant correlations between different modalities. To verify this hypothesis, we conducted several experiments to compare the quality of the extracted basis components using several common algorithms.

The paper is structured in the following way: In the next section we provide a short review of various decomposition approaches, namely k-Means, PCA, NMF and UNMF. Section 3 evaluates the application of these methods to multimodal and high-dimensional data in a turbine blade test scenario. In three studies we compare the abilities of the algorithms at finding local, relevant correlations between modalities and in predicting one modality from another. Finally, section 4 discusses the results and makes some conclusions.

2. Decomposition of Multimodal Data

The starting point of all decomposition approaches is a set of L vectors pooled in an input matrix $\mathbf{X}$:

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_L] \in R^{M \times L}. \quad (1)$$

Each data vector $\mathbf{x}_i$ can be regarded as an observation of M random variables. These random variables might represent the pixels of an image, air pressure, joint angles, temperature or any other numeric information. If the input data consists of several modalities, it is called multimodal and the data from the different modalities are concatenated within the vectors $\mathbf{x}_i$. Decomposition approaches aim for a more compact, approximate representation of this data matrix using a small set of $N < M \ll L$ meaningful components spanning a new vector space, $\mathbf{F}$:

$$\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_N]. \quad (2)$$

This assumes, that the data vectors $\mathbf{x}_i$ are highly correlated and span only a low-dimensional subspace of R^M. In this

case, the basis vectors, $\mathbf{f}_i$, will express typical correlations within the training set, and also correlations between different modalities of the data. The vector space $\mathbf{F}$ is specialized to compactly represent observations similar to the training data. Expressing the data vectors $\mathbf{x}_i$ with respect to this new basis yields an approximation matrix $\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_L]$ and the coefficients of the linear combination – also known as encodings – form an $N \times L$ matrix $\mathbf{G}$. Formally, we can restate this approach as matrix factorization:

$$\mathbf{X} \approx \mathbf{F} \cdot \mathbf{G} \equiv \mathbf{R}, \tag{3}$$

$$\mathbf{G} = \mathbf{F}^{-1} \cdot \mathbf{X}. \tag{4}$$

If $\mathbf{F}$ is non-square or singular, matrix inversion is not possible, but a best fit solution (least squares) can be estimated utilizing Moore-Penrose pseudoinverse. To circumvent negative coefficients enforcing a parts-based representation, Non-Negative Least Squares (25) should be used to determine the encodings. The accuracy of the representation based on the basis $\mathbf{F}$ is measured by the reconstruction error between the original data, $\mathbf{X}$, and its approximation, $\mathbf{R}$:

$$e = \|\mathbf{X} - \mathbf{F} \cdot \mathbf{G}\|_2 = \|\mathbf{X} - \mathbf{R}\|_2 . \tag{5}$$

Different algorithms are distinguished by the restrictions imposed on $\mathbf{F}$ and $\mathbf{G}$. For example, NMF methods (14) constrain both basis vectors and encodings to non-negative values to avoid complex cancellations of features and facilitate their interpretability. Other approaches enforce sparseness (12)(17) or orthogonality (7)(4) constraints. The examined algorithms are especially effective with uni-modal data distributions. Even if the data is composed from several modalities, their distribution is not necessarily multi-modal, i.e., having multiple peaks. For the consideration of multi-modal distributions please see alternatives such as Canonical Correlation (11) or Linear Discriminant Analysis (9).

2.1 k-Means Clustering

k-Means clustering (18) is not a classic decomposition approach. Nevertheless, it represents observations $\mathbf{x}_j$ by a set of k prototypes $\mathbf{f}_i$ resulting in an encoding, which is extremely sparse: only the coefficient associated to the nearest prototype is equal to one. The prototypes are obtained in an iterative process minimizing the total intra-class variance:

$$\min V_{total} = \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} (\mathbf{x}_j - \mathbf{f}_i)^2. \tag{6}$$

k-Means clustering does not necessarily result in optimal partitions. Multiple executions of the algorithm or more sophisticated initialization approaches like in k-Means++ (2) reduce sensitivity to the initial randomly selected cluster seed points. Another issue of k-Means is the determination of an adequate cluster number k. Coarsening and refinement are two techniques that adapt k during clustering (18). Relaxed forms of k-Means clustering allow associations to several classes simultaneously and have close ties to PCA and NMF (5) (6) (16). In fact, Uni-orthogonal NMF has been shown to be equivalent to k-Means clustering under the condition $\mathbf{G}^T \cdot \mathbf{G} = \mathbf{I}$ (7). Some important applications of k-Means are Voronoi tessellation (13), image segmentation (23), unsupervised classification (18) and initialization of NMF-based decomposition algorithms (7).

2.2 Principal Component Analysis (PCA)

PCA (22) is a method of multivariate statistics that is closely related to Singular Value Decomposition (SVD) (26). Successful applications of PCA are found in many fields, e.g., face recognition (24), gene expression analysis (26) or its use as a preprocessing whitening step before application of Independent Component Analysis (ICA) (3).

PCA tries to compute a new subspace, $\mathbf{F}$, whose basis vectors, $\mathbf{f}_i$, are pairwise orthogonal and thereby form an orthogonal set. These principal components point in the directions of the largest variances, σ_i^2. A projection into this vector space decorrelates the input data and increases its statistical dispersion. One analytical approach to identify the principal components is based on the eigenvalue decomposition of the zero-mean input data covariance matrix $\mathbf{C}$ (24).

2.3 Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF) (14) (15) is based on the idea that for perception of the whole perception of its parts is needed. With this in mind, the algorithm was designed to learn parts-based representations. This is in contrast to PCA and vector quantization algorithms like k-Means, which learn holistic representations. Due to its strict limitation to non-negative values, only additive combinations of features are possible, preventing cancellation among elements during superposition. Lee and Seung (15) propose multiplicative update rules for matrices $\mathbf{F}$ and $\mathbf{G}$ to iteratively minimize either the squared error of approximation Eq. (7) or the Kullback-Leibler divergence Eq. (8). Several variations of standard NMF have been derived to further enhance sparseness (4)(8)(12)(17), orthogonality (4)(7) or simultaneous clustering of rows and columns (7).

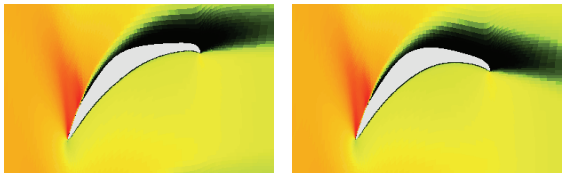
$$\min_{\mathbf{F}\geq 0, \mathbf{G}\geq 0} \|\mathbf{X} - \mathbf{F} \cdot \mathbf{G}\|_2 \qquad \text{(square error)} \tag{7}$$

$$\min_{\mathbf{F}\geq 0, \mathbf{G}\geq 0} D\left(\mathbf{X} \,||\, \mathbf{F} \cdot \mathbf{G}\right) \qquad \text{(Kullback-Leibler)} \tag{8}$$

NMF does not always produce sparse, parts-based representations. To improve the interpretability of the decomposition results UNMF (7) imposes an additional constraint leading to the orthonormalization of either $\mathbf{F}$ or $\mathbf{G}$. A combination of the two uni-orthogonal variants is used in Tri-Factorization (7) to perform a bi-orthogonal 3-factor NMF.

3. Application to a Turbine Blade Data Set

To investigate the practical abilities of the presented decomposition approaches at finding local inter-modal correlations, we compare these methods exemplarily in a test scenario using turbine blade shapes, which are of similar complexity to the ones used in an evolutionary optimization framework in (19). The input to the algorithms consists of 2D blade profiles – represented as binary images – and their associated aerodynamic flow fields – represented by pressure values calculated from a simulated draft (see Fig. 1). To simulate the fluid dynamic properties of the blade designs



(a) upper-tail blade (b) lower-tail blade

Figure 1: The training data of the first study is limited to variations of two basic blade shapes differing in the trailing edges. The associated aerodynamical flow fields are represented by pressure values and were calculated using CFD to simulate a draft.

Figure 2: Color map used to visualize pressure data. Values near zero are represented by black color, extreme negative or positive values by blue and red respectively.

we used an in-house Navier-Stokes flow solver (1). The set of turbine blades comprises shapes that are not necessarily suitable for practical application and are thus only of academic interest. Nevertheless, the performed studies show characteristical properties of the evaluated algorithms. We conducted three studies to compare the decomposition approaches based on their efficiency at finding local, relevant correlations and their ability to predict one modality from another. To visualize the pressure data we utilize an adjusted jet color map shown in Fig. 2. It maps small absolute values to black in order to show their marginal influence in linear combinations. Larger positive and negative numbers are mapped onto ranges of red to blue colors respectively. To enhance perception of the inter-modal correlations, shapes are depicted on top of the according flow fields using shades of gray. Apart from the components of PCA, there is insignificant overlap of form and pressure in the images. The rows of the depicted encoding matrices show the activation of the basis components for each training sample. The selection of basis components is not unique for all decomposition approaches, except for PCA. Any positive monomial matrix $\mathbf{P}$ that produces a (positive) scaling and permutation of basis components results in another decomposition having identical approximation quality:

$$\mathbf{X} \approx \mathbf{F} \cdot \mathbf{G} = \mathbf{F}\mathbf{P} \cdot \mathbf{P}^{-1}\mathbf{G} = \tilde{\mathbf{F}} \cdot \tilde{\mathbf{G}}\,. \tag{9}$$

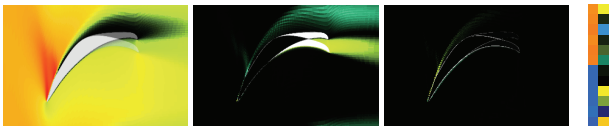
In order to rank the basis components, $\mathbf{f}_i$, according to their contribution to a reduction of the reconstruction error, we sort them with respect to the expression $\|\mathbf{g}_i^T\| \cdot \|\mathbf{f}_i\|$ where $\mathbf{g}_i^T$ is the i-th row vector of $\mathbf{G}$ corresponding to the average contribution of $\mathbf{f}_i$ to an approximation of the training set.

3.1 Study 1: Identifying Two Basic Blade Shapes

To verify the ability of the decomposition algorithms to identify common constituents of a data set, the first study considers a simple toy data set consisting of two basic blade shapes as shown in Fig. 1. Both shapes have a common front

(a) Basis components 1 to 3. (b)

Figure 3: k-Means generates holistic, prototypical basis components (a) with maximal sparseness in the orthogonal encodings (b). Only one class is activated at a time.



(a) Mean vector and first two principal components. (b)

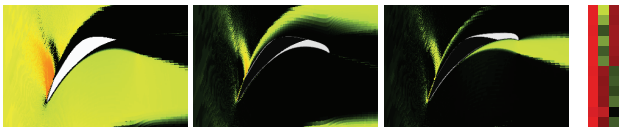
Figure 4: PCA extracts holistic, orthogonal basis components (a). Principle components and encodings (b) have arbitrary signs. Superposition of positive and negative parts.

part and only differ in their trailing edges. The data set contains six small variations of each basic shape. We always compute three basis components, although two should suffice for this input data. Hence, the third component should have much less influence on the reconstruction error than the first two. In general, the necessary number of basis components for a data set can be approximated by analyzing the data variances using PCA eigenvalues.

k-Means clustering (see Fig. 3) produces prototypical basis vectors that are holistic and non-negative (due to the non-negativity of the training data). Each basic blade form is represented by a single basis component. The third cluster centroid represents a training sample randomly belonging to either of the basic shapes depending on initialization. The basis vectors are neither orthogonal nor sparse. However, the encodings generated by k-Means are sparse and orthogonal in an extreme way. For every training observation only one class is activated at a time. As depicted in Fig. 4, PCA extracts the mean vector of the training data and the two principle components related to the highest eigenvalues. All three features are holistic and the two principle components are orthogonal. The values of the encodings and the basis vectors have arbitrary signs. The mean vector and the first principle component are able to recreate the two basic blade shapes by summation or subtraction resulting in cancellations between positive and negative parts. The second principle component contributes to a further reduction of the reconstruction error. NMF (see Fig. 5) is able to produce sparse basis components or sparse encodings. If the sparseness of one increases, the sparseness of the other decreases (21). Due to the training data presented in the first study, NMF generates holistic basis components and sparse encodings. These results take several thousand iterations of the algorithm to manifest themselves. Before convergence the basis components are quite similar to the results of UNMF. The reason for this change in strategy from sparse to holis-

(a) Basis components 1 to 3. (b)

Figure 5: NMF decomposition results: holistic, non-sparse basis components (a) and sparse encodings (b) caused by nature of training data, limited to non-negative values.



(a) Basis components 1 to 3. (b)

Figure 6: Parts-based UNMF decomposition results: sparse, orthogonal basis components (a), non-sparse encodings (b), limited to non-negative values.

tic basis vectors is rooted in a reduced reconstruction error for this specific data set that consists only of small variations of two basic blade shapes. All values of the encodings and the basis components are non-negative. The first basic blade form is approximated by two similar basis vectors. The second basic blade form is represented by the remaining component. Figure 6 confirms that UNMF with an orthogonality constraint imposed on the basis vectors calculates sparse, pairwise orthogonal basis vectors and sparse encodings. The basic blade shapes are constructed from a common front part and a higher or lower tail. Correlations between the two modalities, shape and aerodynamical flow field, can be seen clearly in Fig. 6. A noticeable detail is that the flow field area above the front, upper surface of the blade shape is independently represented by the second and third basis components, instead of employing the common first feature. Closer inspection reveals that indeed this pressure area slightly differs for both blade shapes.

3.2 Study 2: No Shape Restrictions

In the second study, we appraise functional components that emerge from turbine blades without any form restriction. In order to cover a higher variance of blade shapes we generate 250 diverse observations consisting of both modalities, shape and pressure. The turbine blades are deformed and thus only of academic interest. Figure 7 shows some of the training samples. We extract subspaces spanned by 5, 10, 20, 35 or 49 basis components. The variation of the number of basis vectors allows us to analyze the reconstruction errors more detailed. Additionally, we form a test data set comprising 65 different observations to measure the ability to generalize and represent novel though similar data.

The results of k-Means clustering are very similar to those of the first study. The basis components are prototypical and the encodings maximal sparse. The ability to generalize is limited to the selection of the nearest cluster centroid. The

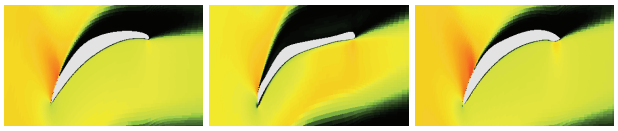
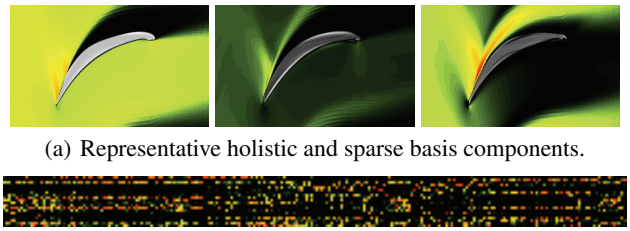


Figure 7: Some observations of the highly variable training data used in study 2. No shape restrictions are imposed.



(a) Representative holistic and sparse basis components.

(b) Sparse encodings matrix.

Figure 8: Three out of 20 basis components (a) calculated by NMF on 250 observations and encodings (b).

holistic principal components of PCA accomplish very small reconstruction errors, but are difficult to interpret due to positive and negative values. The orthogonal basis components of UNMF have a big variance in locality. Some features mainly represent inner parts of blade shapes without any defining influence on the flow fields. This leads to zero activation in the associated pressure modality. Other basis vectors have significant activation in over 30 % of the pressure pixels. The increased sparseness in some basis components, especially in the pressure modality, could be the reason for the poor reconstruction of the training data set. NMF shows fundamental changes in comparison to the first study. For this data set most basis components and the encoding matrix are sparse. Figure 8 depicts three representative basis vectors and the encodings calculated by NMF. The first illustrated component represents a holistic basic shape and its associated flow field. The second feature shows an interesting correlation between the lower, frontal surface and the pressure profile above the turbine blade and the third image displays a basis vector focusing mainly on a single modality, the pressure profile. All NMF runs show qualitatively similar components, independent of their number.

Figure 9 illustrates the approximation of one training sample by NMF. The three most activated basis components are those visualized in Fig. 8. The absolute difference between the original data and its reconstruction highlights the areas that could not be represented adequately. As can be seen from Tab. 1(a) and 1(b), the normalized mean squared reconstruction error,

$$\text{NMSE} = \frac{1}{L}\sum_{i=1}^{L}\frac{\|\mathbf{x}_i - \mathbf{r}_i\|^2}{\|\mathbf{x}_i\|^2}, \tag{10}$$

decreases for all methods uniformly, if the number of available basis components (BC) is increased from 5 to 49. Furthermore, it is shown that the subspace trained on the observed data is able to represent the novel though similar test

(a) Original, encodings, reconstruction, absolute difference.

Figure 9: Reconstruction of a single training observation. Compare Fig. 8 for the three most activated features.

(a) Original, encodings, reconstruction, absolute difference.

Figure 10: Encodings calculated only on pressure and then used to reconstruct the missing shape modality.

data set. However, in this generalization test the reconstruction errors are slightly higher. We also evaluated the impact of multimodality on the reconstruction error. To this end, Tab. 1(c) compares the error values obtained from separate decomposition of both modalities and subsequent accumulation of the individual errors (column "Sep.") with those reconstruction errors obtained from joint decomposition of both modalities (column "Com."). If the basis components need to represent several modalities, the length of the observation vectors increases, which makes it more difficult to identify similarities and extract features. This leads to slightly higher reconstruction errors as a trade-off for the learned correlations. Regarding the reconstruction error, PCA performs best. It is closely followed by NMF, whose relative generalization ability is unmatched. UNMF is worse than k-Means in approximation of the observed training data due to the unused pressure modality in some features, but it outperforms the clustering method in reconstruction of the novel test data set. For all evaluated algorithms we find that the multimodality of the data slightly degrades reconstruction performance. Additional constraints such as non-negativity or orthogonality further increase the reconstruction error. Nevertheless, these factors render it possible to extract more *meaningful* features and to find *local* correlations between modalities, which increase the quality of the basis vectors, i.e., their relevance and interpretability.

3.3 Study 3: Reconstruction of Missing Modalities

In the final study, we focus on the ability to reconstruct missing modalities based on the inherent correlations between the components' subparts. This means that the subspaces comprising 49 basis vectors computed for the second study on 250 observations are used to represent novel test data consisting of only shape or pressure information. The basis vectors are specialized to represent turbine blades. To project the test observations into the spanned subspace, we utilize the Moore-Penrose pseudoinverse. The procedure might introduce some negative encodings, because the least

Approach	(a) Study 2 Observed Data					(b) Study 2 Test Data				
	5	10	20	35	49	5	10	20	35	49
K-Means	52	36	29	23	21	87	88	78	78	77
PCA	38	28	19	13	10	74	53	44	38	34
NMF	40	31	22	17	15	74	59	48	38	36
UNMF	54	46	45	40	40	82	71	70	59	53

Approach	(c) Study 2 Observed Data		Test Data		(d) Study 3 Test Data	
	Sep.	Com.	Sep.	Com.	PR	SR
K-Means	19	21	70	77	80	104
PCA	9	10	32	34	38	57
NMF	13	15	32	36	51	71
UNMF	24	40	40	53	84	156

Table 1: Study 2: Normalized mean squared reconstruction errors (NMSE) for observed (a) and test data set (b) against number of basis components (BC). (c): 49 BC, Sep.: separate decomposition of modalities, Com.: combined decomposition of modalities. Study 3: Reconstruction of missing modalities (d), PR: pressure reconstructed, SR: shape reconstructed. Values scaled by factor 10^3.

squares solution for the linear combination is not restricted to non-negative contributions. However, this does not prevent the estimation of missing modalities, when we use the encodings gained from the first step to reconstruct all modalities, including the previously neglected ones. If the test data set is similar to the training observations used to calculate the features, the missing information is approximated successfully. Figure 10 depicts the reconstruction of a missing shape modality for a novel test observation. The reconstruction errors for the approximation of missing modalities can be seen in Tab. 1, section (d). "PR" stands for pressure reconstructed. This means that the encodings were estimated solely on the shape information and then used to reconstruct form and pressure. "SR" describes the counterpart: shape reconstructed. The encodings are calculated using only the flow fields and then utilized to estimate both modalities, including the former missing shape. In this study PCA excels due to its strong capability to find correlations. NMF performs second best, representing a good compromise between interpretability and reconstruction quality. UNMF outperforms k-Means in reconstructing aerodynamical flow fields, due to k-Means' lack of flexibility caused by its limitation to only select the nearest cluster mean. On the other hand, UNMF struggles approximating missing shapes, because of some basis components having "empty" pressure modalities, which are not suited to calculate the needed encodings. In this case, even k-Means results in smaller reconstruction errors. Nevertheless, the decomposition algorithms are able to approximate missing modalities successfully within a sufficiently small error margin. A practical application could be the generation of new design concepts by combining different local contributions to form a desired global flow field and reconstructing the associated shape primitives accordingly. If the result is too inaccurate, the

number of basis components could be increased or the training data adapted. The calculated form could be a starting point for further analysis by an engineer.

4. Conclusion

In this paper, we have proposed the application of standard decomposition approaches to find local correlations in multimodal data. In a test scenario, we applied these methods to correlate the local shape of turbine blades with their associated aerodynamic flow fields. We compared the decomposition algorithms with regards to their efficiency at identifying local, relevant correlations and their ability to predict one modality from the other. For the evaluated algorithms we found that the multimodality of the data and additional constraints such as non-negativity or orthogonality can improve the algorithm's ability to extract meaningful basis components with a trade-off of a small decrease in the reconstruction performance. With regards to the reconstruction error, PCA performs best, because it decorrelates the input data and increases its statistical dispersion. Both NMF-based approaches revealed an interesting inter-modal correlation between a shape feature on the lower side of the blades and the aerodynamical flow field on their upper side. Furthermore, it was possible to predict one modality from the other, which is evidence for the learned inter-modal correlation and might turn out to be useful in design optimization processes. New design concepts could be generated by combining different local contributions to form a desired global flow field and selecting the associated shape primitives accordingly. Also, interchangeable local features might be identified that have similar influence on the air flow but have different shapes.

5. Acknowledgments

Daniel Dornbusch gratefully acknowledges the financial support from Honda Research Institute Europe for the project "Generation of a Generic Decomposition Framework for Affordance-based Grasping and Design Optimization."
bibliographystyleaaai

References

Arima, T.; Sonoda, T.; Shirotori, M.; Tamura, A.; and Kikuchi, K. 1999. A numerical investigation of transonic axial compressor rotor flow using a low-reynolds-number k-e turbulence model. *ASME Journal of Turbomachinery* 121(1):44–58.

Arthur, D., and Vassilvitskii, S. 2007. k-Means++ The advantages of careful seeding. In *Proc. Symposium on Discrete Algorithms (SODA '07)*, 1027–1035.

Bartlett, M. S. 1998. *Face image analysis by unsupervised learning and redundancy reduction*. Ph.D. Dissertation, University of California, San Diego.

Bozakov, Z.; Graening, L.; Hasler, S.; Wersing, H.; and Menzel, S. 2008. Unsupervised extraction of design components for a 3d parts-based representation. In *Proc. IEEE IJCNN '08*, 2009–2016.

Ding, C.; He, X.; and Simon, H. D. 2005. On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proc. SIAM Int. Conf. on Data Mining (SDM '05)*, 606–610.

Ding, C.; Li, T.; and Jordan, M. I. 2006. Convex and semi-nonnegative matrix factorizations. Technical report, Lawrence Berkeley National Labratory.

Ding, C. 2006. Orthogonal nonnegative matrix tri-factorizations for clustering. In *Proc. Int. Conf. on Knowledge Discovery and Data Mining (KDD '06)*, 126–135.

Eggert, J., and Körner, E. 2004. Sparse coding and NMF. In *Proc. IEEE Int. Joint Conf. on Neural Networks*, 2529–2533.

Fisher, R. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7:179–188.

Hasler, S.; Wersing, H.; and Körner, E. 2007. Combining reconstruction and discrimination with class-specific sparse coding. *Neural Computation* 19(7):1897–1918.

Hotelling, H. 2009. The most predictable criterion. *Journal of Educational Psychology* 26:139–142.

Hoyer, P. O. 2004. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research (JMLR '04)* 5:1457–1469.

Inaba, M.; Katoh, N.; and Imai, H. 1994. Applications of weighted voronoi diagrams and randomization to variance-based k-clustering. In *Proc. 10th ACM Symposium on Computational Geometry*, 332–339.

Lee, D. D., and Seung, H. S. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755):788–791.

Lee, D. D., and Seung, H. S. 2000. Algorithms for non-negative matrix factorization. In *Proc. Int. Conf. on Neural Information Processing Systems (NIPS)*, 556–562.

Li, T., and Ding, C. 2006. The relationships among various nonnegative matrix factorization methods for clustering. In *Proc. 6th Int. Conf. on Data Mining (ICDM '06)*, 362–371.

Li, S. Z.; Hou, X. W.; Zhang, H. J.; and Cheng, Q. S. 2001. Learning spatially localized, parts-based representation. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, 207–212.

MacQueen, J. B. 1967. Some methods for classification and analysis of multivariate observations. In *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability (BSMSP '67)*, 281–297. University of California Press.

Menzel, S., and Sendhoff, B. 2008. Representing the change - free form deformation for evolutionary design optimisation. In Yu, T.; Davis, D.; Baydar, C.; and Roy, R., eds., *Evolutionary Computation in Practice*. Springer. chapter 4, 63–86.

Osowski, S.; Majkowski, A.; and Cichocki, A. 1997. Robust PCA neural networks for random noise reduction of the data. In *Proc. of the 1997 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP '97)*, 3397–3400. Washington, DC, USA: IEEE Computer Society.

Pascual-Montano, A.; Carazo, J. M.; Kochi, K.; Lehmann, D.; and Pascual-Marqui, R. D. 2006. Nonsmooth nonnegative matrix factorization (nsNMF). *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 28(3):403–415.

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2(6):559–572.

Shapiro, L. G., and Stockman, G. C. 2001. *Computer Vision*. Prentice Hall.

Turk, M., and Pentland, A. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience (JCN '91)* 3(1):71–86.

van Benthem, M. H., and Keenan, M. R. 2004. Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems. *Journal of Chemometrics* 18(10):441–450.

Wall, M. E.; Rechtsteiner, A.; and Rocha, L. M. 2003. Singular value decomposition and principal component analysis. In *A Practical Approach to Microarray Data Analysis*. Kluwer. chapter 5, 91–109.