Knowledge Extraction from Unstructured Surface Meshes

Lars Gräning, Markus Olhofer, Bernhard Sendhoff

2007

Preprint:

This is an accepted article published in Proceedings of the 8th International Conference on Intelligent Data Egineering and Automated Learning (IDEAL). The final authenticated version is available online at: https://doi.org/[DOI not available]

Knowledge Extraction from Unstructured Surface Meshes

Lars Graening, Markus Olhofer and Bernhard Sendhoff

Honda Research Institute Europe GmbH,
Carl-Legien-Strasse 30, D-63073 Offenbach/Main, Germany
{lars.graening,markus.olhofer,bernhard.sendhoff}@honda-ri.de

Abstract. We propose methods that allow the investigation of local modifications of aerodynamic design data represented by discrete unstructured surface meshes. A displacement measure is suggested to evaluate local differences between the shapes. The displacement measure provides information on the amount and direction of surface modifications. Using the displacement measure in conjunction with statistical methods or data mining techniques provides meaningfull knowledge from the data set for guiding further shape optimization processes.

1 Introduction

In the field of 3D aerodynamic shape optimization, a large amount of geometric and flow field data is generated during the design process that usually encompasses several optimization runs, manual design phases and experiments. Typically, only the most promising results with regard to one or more possibly competing performance indices are exploited to define the overal result of the design process. However, a lot of information that could be condensed into comprehensive rules or observations concerning the design process in general is hidden in all of the data. Even poorly performing shapes can provide interesting insight into the fluid-dynamics of the problem and into the dynamics of the search process. This knowledge extracted from the large amount of data can be prepared in such a way that it can be used by the engineer or by an follow-up computational design and optimization processse. This type of knowledge extraction is the major focus of the present paper.

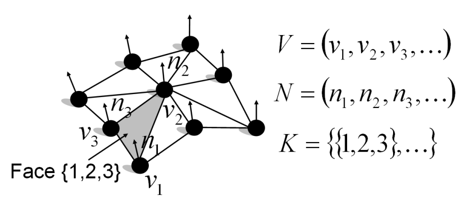
Obayashi et al. [10] were one of the first who addressed the problem of knowledge extraction from existing design data in order to gain some insights into the complex relationship between geometry and performance measurements. They used self organizing maps (SOM) in order to find groups of similar designs and for multicriteria performance improvements and tradeoffs. Although their methods have been applied to super sonic wing design, the data and design parameter sets were small, uniform and well defined. If different optimization runs have been performed with different design parameters, one first has to find an adequate representation which captures all shape variations and which can be applied to various data mining techniques.

Therefore, we suggest the use of unstructured surface meshes as a general representation for analyzing a given set of designs resulting from different shape optimization runs. Each optimization can be a manual or a computational process and can be based on different shape descriptors. The unstructured surface mesh as a general representation allows the analysis of local shape modifications and their influence on the performance value(s). In this paper, we will propose a displacement measure between surface meshes. The combination of the displacement measure with techniques from statistics and data mining allows the extraction of useful knowledge from the design database for the support of further optimization processes.

The paper is organized as follows. In Section 2, we will introduce unstructured surface meshes and outline the new displacement measure in Section 3, which captures local differences between designs. How we can extract knowledge from the displacement and performance data is described in Section 4. In the last section, we summarize the paper.

2 Surface Representation

For the optimization of the shape of three dimensional geometries often different parametric representations are used [2] which makes it difficult or even impossible to analyze the whole data set based on the applied parameterization. Therefore, we suggest unstructured triangular surface meshes as a general representation to describe the surface of each design. Most shape representations can be converted to unstructured surface meshes, see e.g. [3], [4].



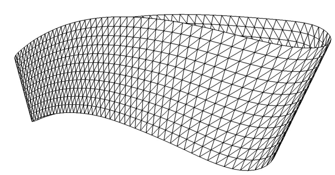


Fig. 1. Illustration of a) the specification of a triangular surface mesh $\mathcal{M} : (\mathcal{V}, \mathcal{N}, \mathcal{K})$ and b) an example of a triangular surface mesh for a 3D turbine blade.

For the description of the surface mesh we start with the mathematical framework given in [5]. It is assumed that the shape of a 3D design is described using a polygonal surface mesh $\mathcal{M}$, which is a partially linear approximation of the contour of the design. We postulate that each mesh $\mathcal{M}$ consists of a list of vertices $\mathcal{V}$, a complex $\mathcal{K}$ and a list of normal vectors $\mathcal{N}$. The vertex list $\mathcal{V} = (\boldsymbol{v}_1, ..., \boldsymbol{v}_n)^T$ describes the geometric position of the vertices in $\mathbb{R}^3$, $\boldsymbol{v}_i = (x_1, x_2, x_3)^T$. A vertex can be seen as a sample point of the contour of the design. Each face of the polygonal surface mesh is defined by simplices of the form $\{i_1, i_2, i_3, ..., i_\mu\}$ where $i_l, l \in [1..n]$ are indices pointing to vertices that enclose the polygonal face.

Figure 1 a) illustrates a triangular surface mesh where the number of vertices, which are used to form each polygon, is set to 3. In addition to the vertex list, a list of normal vectors $\mathcal{N} = (\boldsymbol{n}_1, ..., \boldsymbol{n}_n)$ is given. Each normal vector $\boldsymbol{n}_i$ has a defined direction perpendicular to the surface mesh and provides local curvature information at the position of vertex $\boldsymbol{v}_i$. Figure 1 b) illustrates an example of a triangular surface mesh describing the contour of a 3D turbine blade design [1]. The normal vectors point towards the outside of the closed blade contour.

3 Displacement Measurement

Under the assumption that the surface triangulation results in surface meshes for which the location and the number of vertices is sufficiently precise to capture the characteristic changes of all designs in the given data set, the displacement is measured between each vertex on the reference design and each corresponding vertex on the modified design. In order to measure the displacement between two vertices of different surface meshes, the correspondence problem (which vertex from mesh $\mathcal{M}_r$ corresponds to which vertex from mesh $\mathcal{M}_m$) has to be solved and an appropriate metric has to be found for measuring the amount and the direction of the displacement between both vertices. We will not deal with the correspondence problem in this paper, the interested reader is refered to e.g. [7], [6]. In the following, we assume that two corresponding vertices have been identified.

3.1 Definition

The displacement measure should describe the position of a vertex in reference to another design. One way to capture this information is to use the difference vector $\boldsymbol{s}_{ij} = \boldsymbol{v}_i^r - \boldsymbol{v}_j^m$, which is the difference between vertex i of mesh $\mathcal{M}_r$ and vertex j of mesh $\mathcal{M}_m$. The difference vector clearly captures the correct displacement between both vertices. However, the difference vector is sensitive to possible errors resulting from wrong estimations of the corresponding points or from different sampling methods of the surfaces of the geometries. Furthermore, the difference vector requires $d = 3$ parameters for describing the displacement of one vertex in $\mathbb{R}^3$. Thus, to capture the displacement between two complete surface meshes the number of parameters is $3 \cdot n$, where n equals the number of vertices. To overcome the disadvantages of the difference vector, we suggest the following displacement measure:

$$\delta_{i,j}^{r,m} = \delta(\boldsymbol{v}_i^r, \boldsymbol{v}_j^m) = (\boldsymbol{v}_j^m - \boldsymbol{v}_i^r) \circ \boldsymbol{n}_i^r, \delta \in (-\infty, +\infty) \tag{1}$$

The displacement measure is defined as the projection of the difference vector $\boldsymbol{s}_{ij} = (\boldsymbol{v}_i^r - \boldsymbol{v}_j^m)$ onto the normal vector $\boldsymbol{n}_i$ of vertex $\boldsymbol{v}_i$ of the reference design $\mathcal{M}_r$. The absolute value of the displacement measure provides information on the amount of vertex modification while the sign of the displacement measure in conjuction with the normal vector of the vertex provides information on the direction of the vertex modification. The normal vector $\boldsymbol{n}_i$ points towards the normal or positive direction of vertex modification.

3.2 Major Properties

The displacement measure is by definition a vector quantity containing both the magnitude and the direction of vertex modification. If the modified vertex lies above the tangential plane described by the normal vector of the reference vertex, the displacement measure is positive, see Fig. 2 a). Whereas if the vertex lies below the tangential plane (Fig. 2 b) the displacement measure is negative. In the special case when the modified vertex is located within the tangential plane, the displacement measure is zero as shown in Fig. 2 c). If the reference vertex has been modified along the line described by the normal vector, the amount of the displacement measure equals the Euclidean distance between the reference and the modified vertex.

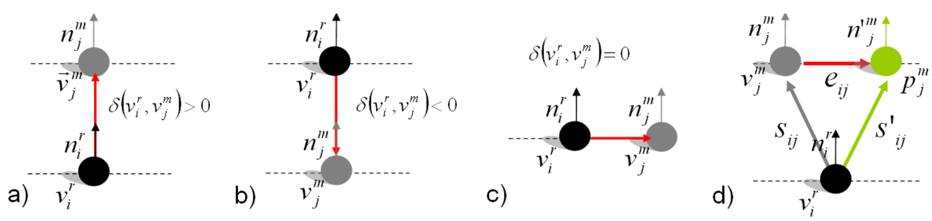


Fig. 2. Examples of the displacement measure. Figures a) and b) illustrate that a vertex displacement parallel (anti-parallel) to the normal direction results in a positive (negative) displacement value. A displacement perpendicular to the normal vector results in a displacement value of zero, as shown in c). Figure d) illustrates the error when calculating the displacement measure, which results from the discretization of the surface and the error when estimating corresponding points.

As Figure 2 d) indicates, the displacement value contains an error, which is mainly the result of the discretization of the surface using triangulation and of the correspondence problem. Formally, this can be written as

$$\delta_{i,j}^{r,m} = (\boldsymbol{s}_{ij} + \boldsymbol{e}_{ij}) \circ \boldsymbol{n}_i^r = \boldsymbol{s}_{ij} \circ \boldsymbol{n}_i^r + \boldsymbol{e}_{ij} \circ \boldsymbol{n}_i^r, \tag{2}$$

where $\boldsymbol{e}_{ij}$ describes the error between the ideal displacement value and the measured displacement value. Under the assumption that the curvature of both surfaces $\mathcal{M}_r$ and $\mathcal{M}_m$ is similar at the position of the corresponding vertices it follows that $\boldsymbol{n}_i^r \approx \boldsymbol{n}_j^m$. Then, the error term from equation 2 can be rewritten as:

$$\boldsymbol{e}_{ij} \circ \boldsymbol{n}_i^r \approx |\boldsymbol{e}_{ij}| cos(\angle(\boldsymbol{e}_{ij}, \boldsymbol{n}_j^m)). \tag{3}$$

If additionally a smooth surface or a small error $|\boldsymbol{e}_{ij}|$ is assumed, $\boldsymbol{e}_{ij}$ is perpendicular to $\boldsymbol{n}_j^m$ and hence $cos(\angle(\boldsymbol{e}_{ij}, \boldsymbol{n}_j^m)) \approx 0$. Thus the error term becomes zero. Therefore, the displacement measure is less sensitive to small errors arising from the surface triangulation or from an incorrect estimation of corresponding points.

Another advantage of the displacement measure compared with the difference vector is that only n parameters are required for the description of the differences between two unstructured surface meshes, where n equals the number of vertices.

4 Knowledge Extraction from Design Data

In aerodynamic design optimization the main goal is to find three dimensional shapes, which are optimal for specific performance measurements like aerodynamic drag or lift under specific constraints, e.g. manufacturing limitations. In general, during the optimization process a large number of shapes are generated and evaluated based on different representations and parameterizations. The result are heterogeneous design data sets from which only a very small number of designs are used in the end to determine the optimal shape (or a set of optimal shapes) which is processed further, e.g. in rapid prototyping devices for experiments. As we noted in the introduction, we aim at exploiting the information contained in the large remaining part of the data set. In this section, we describe how the displacement measure in conjunction with statistical and data mining methods can be used in order to extract meaningfull information (knowledge) from heterogeneous design data sets.

4.1 Displacement Analyzis

Analyzing local modifications in form of vertex displacement helps to gain some insight into the exploration of the design space. Two measures are suggested: the relative mean vertex displacement that provides information on how a vertex has been modified with respect to one reference design and the overall displacement variance that highlights the vertices which have been modified most frequently.

Relative Mean Vertex Displacement In order to get information on local design modifications in reference to one baseline design, we define the *relative mean vertex displacement* :

$$\overline{\delta}_i^r = \frac{1}{N-1} \sum_{m=1, m \neq r}^{N} \delta_{i,j}^{r,m} \tag{4}$$

Given a data set of N unstructured surface meshes $\overline{\delta}_i^r$ evaluates the mean displacement of vertex j of all meshes m from the corresponding vertex i of the reference mesh (baseline) r. The measure provides information on how far a reference vertex has been modified along its normal vector with respect to the whole data set. If $\overline{\delta}_i^r > 0$, the vertex $\boldsymbol{v}_i^m$ has been modified parallel to the normal vector of the vertex and $\overline{\delta}_i^r < 0$ indicates a modification anti-parallel to the normal direction of the vertex. If $\overline{\delta}_i^r = 0$, the vertex has not been modified or the modifications around the reference vertex in the data set have canceled each other out. In order to identify the later situation one can calculate the variance of the deformation values. If there are outliers that affect the calculation of $\overline{\delta}_i^r$, we recommend to use the median instead of the mean in order to retrieve the desired information.

As an example the relative mean vertex displacement has been calculated based on a set of 200 turbine blades from different design optimization runs and a pre-selected references design. For illustration purpose the values have been coded into corresponding color values and mapped onto the surface of the reference blade, Figure 3 a).

Overall Displacement Variance In order to calculate $\overline{\delta}_i^r$, the baseline mesh r must be given. An alternative would be to calculate the mean displacement between all possible shape combinations in the data set. However, this is not sensible, because if the normal vectors of corresponding vertices are similar, it holds that $\delta_{i,j}^{r,m} \approx -\delta_{i,j}^{m,r}$ and as a result such a measure would always tend to zero.

In order to get an overview over the variations of local design modifications an *overall displacement variance* can be defined as follows:

$$\sigma_{\delta_i} = \sqrt{\frac{1}{N(N-1)} \sum_{r=1}^{N} \sum_{m=1, m \neq r}^{N} (\delta_{i,j}^{r,m} - \overline{\delta}_i)^2} \approx \sqrt{\frac{2}{N(N-1)} \cdot \sum_{l=1}^{N} \sum_{m=r+1}^{N} (\delta_{i,j}^{r,m})^2} \tag{5}$$

This measure describes the strength and the frequency of local design modifications based on the whole data set. Following our argument above, we can set $\overline{\delta}_i \approx 0$.

Figure 3 b) shows the overall displacement for the turbine blade data set. In order to visualize high as well as low variances σ_{δ_i} has been displayed in logarithmic scale.

4.2 Sensitivity Analysis

Sensitivity analysis relates the displacement measure to variations of the corresponding performance values.

Relative Vertex Correlation Coefficient The *relative vertex correlation coefficient* R_i^r, see Equation 6, formalizes the linear relation between local modifications in form of vertex displacements and performance values with respect to a chosen reference design. $\phi^{r,m} = f^m - f^r$ is the performance difference between two designs r and m and $\overline{\phi}^r$ is the mean value of the performance differences with respect to the reference design r.

$$R_i^r = \frac{\sum_{m=1, m \neq r}^{N} (\delta_{i,j}^{r,m} - \overline{\delta}_i^r)(\phi^{r,m} - \overline{\phi}^r)}{\sigma_{\delta_i^r} \sigma_{\phi^r}} \tag{6}$$

$$\sigma_{\delta_i^r} = \sqrt{\frac{1}{N-1} \sum_{m=1, m \neq r}^{N} (\delta_{i,j}^{r,m} - \overline{\delta}_i^r)^2}, \qquad \overline{\phi}^r = \frac{1}{N} \sum_{m=1, m \neq r}^{N} \phi^{r,m} \tag{7}$$

$R_i^r > 0$ indicates that moving the vertex parallel to the normal vector is most likely to improve the performance of the design and vice versa. Again two situations can lead to a vanishing R_i^r value. Firstly, the obvious explanation is that an (anti)-parallel modification of the vertex has no effect on the performance measure. Secondly, if the vertex is already located in an optimal position, every modification will reduce the performance and R_i^r will also be close to zero. In order to distinguish between both cases, one could fit a linear model to the displacement and performance difference pairs and calculate the residual of the linear model. This residual provides information on the uncertainty of the correlation coefficient. Of course, the uncertainty of the correlation coefficient might also result from noisy data or non-linear relations between displacement measure and performance differences.

Concerning the 3D turbine blade example the performance of each design is determined by the overall pressure loss of the blade, see [1]. Based on the displacement and pressure loss data the relative vertex correlation has been calculated. The result is illustrated in Figure 3.

Vertex Sensitivity In order to identify vertices that are sensitive to performance changes based on the whole data set without referring to one baseline shape, the Pearson correlation coefficient is calculated based on all pairwise design comparisons. Calculating the mean value for all performance differences obviously results in $\overline{\phi} = 0$. We define the *overall vertex correlation coefficient* as follows (assuming again $\overline{\delta}_i \approx 0$):

$$R_i = \frac{\sum_{r=1}^{N} \sum_{m=1, m \neq r}^{N} \delta_{i,j}^{r,m} \phi^{r,m}}{\sigma_{\delta_i} \sigma_\phi}, \qquad \sigma_\phi = \sqrt{\frac{2}{N(N-1)} \cdot \sum_{r=1}^{N} \sum_{m=r+1}^{N} (\phi^{r,m})^2} \tag{8}$$

The overall vertex correlation coefficient captures the linear relationship between the displacement and performance changes. In order to be less sensitive to outliers or noise in the data, it is reasonable to apply the Spearman rank based coefficient instead of the Pearson correlation coefficient. Since the overall vertex correlation is linear, information is provided to distinguish between those vertices which are more likely to improve the performance by moving them parallel to the direction of the normal vector and those which improve the performance when moving them anti-parallel to the direction of the normal vector.

The overall vertex correlation coefficient has also been calculated for the turbine blade data set. In order to identify most sensitive points a threshold has been applied to the sensitivity values, R_i. The emerged regions can be distinguished into regions of positive and those of negative correlation as shown in Figure 3 d).

In aerodynamic design optimization the interrelation between design modifications and performance changes is often highly non-linear. In order to capture also non-linearities, one can apply information based measures like *mutual information* [11] to determine the sensitivity of vertices. The disadvantage of non-linear methods like mutual information is that the information to predict the direction of design improvement with respect to the normal vector is lost.

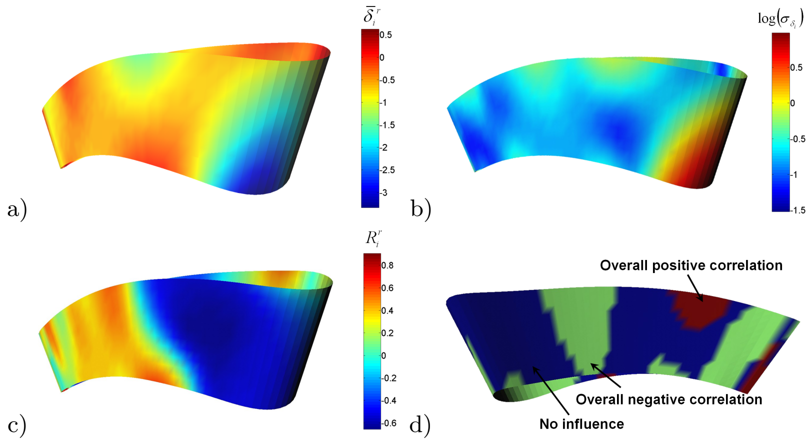


Fig. 3. As an example the suggested methods are calculated based on a data set which consists of 200 3D turbine blades: a) Relative mean vertex displacement, b) Overall displacement variance in logarithmic scale, c) Relative vertex correlation coefficient, d) Sensitive regions of the blade.

4.3 Modeling and Analyzing Interrelations

For the calculation of the measures described above, the displacement of each vertex is considered independent of the others. Especially in aerodynamics the interrelation between distant vertices or design regions and their joint influence on the performance plays an important role. In this section, special characteristics for the extraction of knowledge in form of associative rules based on data from unstructured surface meshes are discussed and illustrated by means of the blade example described above. The rules describe the relation between the displacement of distant vertices and their joint influence on the performance criteria. Modeling the interrelation between input variables is achieved by applying well known modeling techniques like Fuzzy rule induction, Bayesian networks, decision trees or others to the data set, for an overview of techniques see e.g. [8].

Rule Induction Generally, the number of input parameters must be kept small for most modeling techniques in order to produce a small set of interpretable and manageable association rules. With respect to the used shape representation, the number of inputs equals the number of vertices n, which is large in practice. Therefore, a reduction of the number of input parameters is strongly required. Concerning the present turbine blade example, this process consists of the following steps:

1. Perform sensitivity analysis
2. Select most sensitive vertices (e.g. apply threshold to the sensitivity metrics)
3. Cluster sensitive vertices to form sensitive areas (e.g. K-means)
4. Calculate cluster centers of the sensitive areas (the number of input variables depend on the number of clusters)
5. Use displacement of vertices closest to cluster centers for rule extraction

The main task when modeling interrelations between distant design regions or vertices is to extract associative rules which can be interpreted by aerodynamic engineers. These rules are subdivided into relative rules, which refer to a baseline shape, and general rules, which refer to the complete data set.

Besides standard real-valued input for the modeling technique, the input can also be restricted to the sign of the displacement measure. In this case, rules from the two-valued input describe the interrelation between the direction of vertex displacement and the change in the performance value.

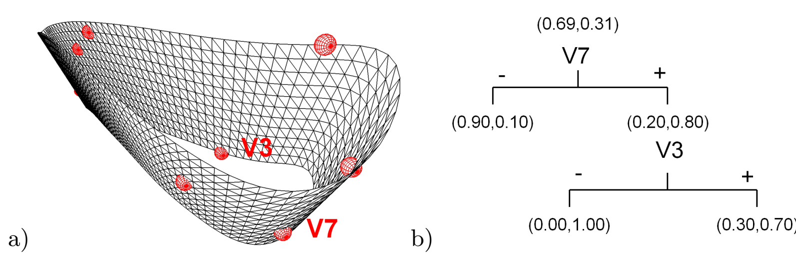


Fig. 4. a) Illustration of the vertices close to the cluster centers of the sensitive areas. b) Simplified decission tree describing the joint interrelation between a subset of vertices and the influence to the preformance (overall pressure loss)

Figure 4 illustrates the reduced subset of parameters (vertices) and a part of the complete classification tree extracted from the turbine blade data set. The classification tree describes the interrelation between the direction of vertex displacement and the change in the overall pressure loss. The rules for the correlated movement of vertices are extracted in form of joint probabilities. For example, moving vertex V7 alone will improve the performance by an probability of $p(\phi > 0|\delta^r_{V7} < 0) = 0.90$. But moving V7 correlated with V3 will always result in an increase in the overall pressure loss, $p(\phi < 0|\delta^r_{V7} > 0, \quad \delta^r_{V3} < 0) = 1.00$.

5 Summary

In this paper, we investigated the problem of how to extract knowledge from the large heterogeneous data sets that usually result from aerodynamic shape optimization processes. Firstly, the aim is to communicate this knowledge to the engineer to increase his/her understanding of the relation between shape and aerodynamic performance, e.g. which part of the design space has been explored and which part has been largely ignored in the past design processes. Secondly, the information from the data set can be used in order to improve the ongoing optimization process, e.g. by specifying parameters of the optimization algorithms or by increasing the generalization capabilities and reducing the approximation erros of surrogate models [13].

The main contribution described in this paper is the formulation of a displacement measure that acts on a generalized shape representation - the unstructured mesh. Based on the displacement measure a number of methods and approaches for displacement analysis, sensitivity analysis and rule extraction were suggested and formulized.

In order to demonstrate the feasibility of the suggested approach, we have shown examples for the proposed measures from a data set taken from the optimization of a 3D turbine blade.

References

1. M. Hasenjäger, B. Sendhoff, T. Sonoda and T. Arima, Three dimensional aerodynamic optimization for an ultra-low aspect ratio transonic turbine stator blade, *Proceedings of the ASME Turbo Expo*, 2005.
2. J. Samareh, A Survey of Shape Parametrization Techniques, *Intern. Forum on Aeroelasticity and Structural Dynamics Conf.*, 1999
3. Y. Ito and K. Nakahashi, Surface Triangulation for Polygonal Models Based on CAD Data, *Intern. Journal for Numerical Methods in Fluids*, vol. 39, 2002
4. K. Shimada, Anisotropic Triangular Meshing of Parametric Surfaces via Close Packing of Ellipsoidal Bubbles, *Intern. Journal of Computational Geometries and Applications*, vol. 10, no. 4, pp. 400–424, 2000
5. M. Alexa, Recent Advances in Mesh Morphing, *Computer Graphics forum*, vol. 21,no. 2, pp. 173–197, 2002
6. V. Jain and H. Zhang, Robust 3D Shape Correspondence in the Spectral Domain, *IEEE Intern. Conf. on Shape Modeling and Applications, SMI'06*, pp. 19, 2006
7. Y. Wang, B. S. Peterson and L. H. Staib, Shape-Based 3D Surface Correspondence Using Geodesics and Local Geometry, *Computer Vision and Pattern Recognition*, pp. 644–651, 2000
8. *The Data Mining and Knowledge Discovery Handbook*, Edited by Oded Maimon and Lior Rokach, Springer, 2005
9. L. Breiman,J.H. Friedman, R.A.Olshen and C.J. Stone, *Classification and Regression Trees* ,CA.: Wadsworth, Inc., 1984
10. S. Obayashi and D. Sasaki, Visualization and Data Mining of Pareto Solutions Using Self-Organizing Map, *In Proc. Second Intern. Conf. on Evolutionary Multi-Criterion Optimization*, EMO 2003, vol. 2632/2003, pp. 71, Faro, Portugal, 2003

11. R. Steuer, J. Kurths, C.O. Daub, J.Weise and J.Selbig, The Mutual Information: Detecting and Evaluating Dependencies between Variables, *Bioinformatics*, 2002
12. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1988
13. Y. Jin, A Comprehensive Survey of Fitness Approximation in Evolutionary Computation, *Soft Computing*, vol. 9, no. 1, pp. 3–12, 2005