Towards Automatic Feature Vector Optimization for Multimedia Applications

Tobias Schreck
TU Darmstadt
Fraunhoferstrasse 5
Darmstadt, Germany

tschreck@gris.informatik.tu-
darmstadt.de

Dieter Fellner
Fraunhofer IGD and TU
Darmstadt
Fraunhoferstrasse 5
Darmstadt, Germany

d.fellner@igd.fraunhofer.de

Daniel Keim
University of Konstanz
Universitaetsstr. 10
Konstanz, Germany

keim@inf.uni-konstanz.de

ABSTRACT

We systematically evaluate a recently proposed method for unsupervised discrimination power analysis for feature selection and optimization in multimedia applications. A series of experiments using real and synthetic benchmark data is conducted, the results of which indicate the suitability of the method for unsupervised feature selection and optimization. We present an approach for generating synthetic feature spaces of varying discrimination power, modeling main characteristics from real world feature vector extractors. A simple, yet powerful visualization is used to communicate the results of the automatic analysis to the user.

Categories and Subject Descriptors

H.2.4 [**Information Systems**]: Multimedia Databases; I.5.2 [**Pattern Recognition**]: Feature evaluation and selection

Keywords

Feature vectors, discrimination power, feature selection, Self-organizing map.

1. INTRODUCTION AND BACKGROUND

Many multimedia applications need to determine the similarity between multimedia objects. E.g., in content-based retrieval, cluster analysis, and fingerprinting applications, discriminating similarity functions are required. One of the most popular approaches for implementing similarity functions for multimedia data is to capture important features of the underlying objects by numeric vectors, referred to as feature vectors (FVs) [5]. Distances between FVs are computed using a suitable FV metric, and associated with the degree of similarity between the objects represented by FVs.

For prominent multimedia data types such as images, music pieces, video shots, or 3D models, an abundance of different similarity functions based on FVs have been proposed to date. For *discriminating* (or effective) FV representations,

distances measured in FV space closely resemble similarity relationships in object space. The design of discriminating FV extractors for multimedia data is, however, a difficult problem, which to date is solved mostly by a heuristic approach: Candidate FVs are composed by selected numeric features, and experimentally benchmarked using preclassified (supervised) data. The main problem with this approach is that it relies on the availability of supervised information, which (a) is expensive to obtain, (b) may be data dependent and not generalize to different applications. Furthermore, supervised benchmark outcomes may be unstable regarding benchmark design choices [8].

Methods for the *unsupervised* estimation of FV discrimination power, allowing automatic benchmarking of different FV configurations, are therefore desirable. A number of theoretical approaches proposed for unsupervised FV analysis [1, 6] to the best of our knowledge have not been practically exploited yet. In [9], we addressed the problem of unsupervised discrimination estimation by means of an analysis function defined over a compressed (clustered) representation of the *unclassified* feature vector data. The analysis function serves as an estimator for the discrimination power one may expect in a given *feature space*, understood as the embedding of a given data set in a metric space using a given feature extractor. We here complement the work introduced in [9] by an in-depth evaluation presenting a systematic series of experiments using real world and synthetic data. The study shows the robustness of the unsupervised feature selection technique, demonstrating its potential as an unsupervised complement or alternative to the long, iterative, and costly process of supervised optimization of candidate feature vector extractors.

Section 2 recalls the basic idea behind the unsupervised discrimination power estimator. Section 3 applies the estimator on a real world data set, evaluating its performance for *selecting feature vectors* from a set of competing vectors, and for *selecting the optimal dimensionality* for a given feature vector. Both selection problems frequently occur when designing multimedia applications which rely on feature vectors to discriminate objects. In Section 4, the estimator is evaluated on a synthetic data set, modeling feature spaces of varying discrimination power. Both evaluations are illustrated using visualization. Section 5 concludes the effectiveness of the estimator, and outlines future work.

2. UNSUPERVISED FEATURE SELECTION

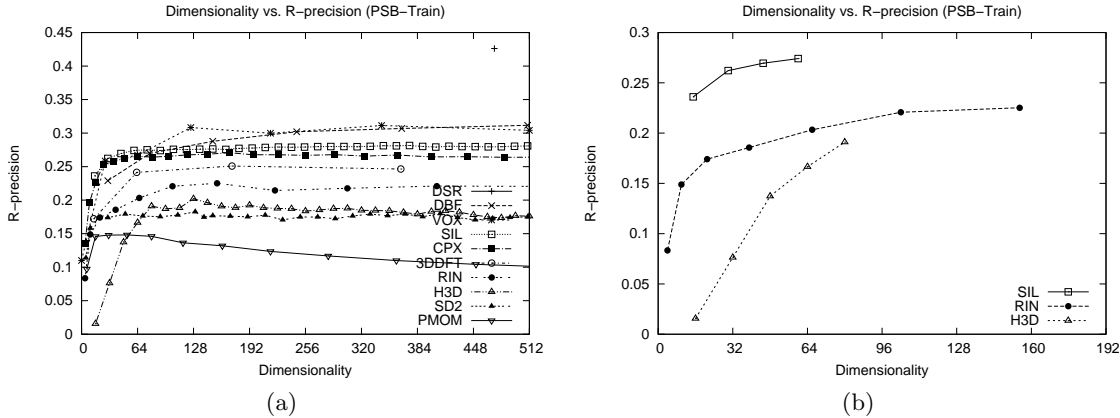In [9], we proposed an unsupervised estimator of FV space

Figure 1: Baseline supervised benchmark against which the unsupervised analysis method will be evaluated. Chart (a) plots dimensionality vs. precision of the ten 3D feature spaces measured on the PSB data set. Chart (b) shows a subset of chart (a) used in the experiments discussed in Section 3.2.

discrimination power based on the following hypothesis: *Discrimination power provided in a given FV space can be estimated by the uniformity of the histogram of distances between the clusters in the respective FV space.* [9] More specifically, the hypothesis states that the discrimination power contained in a given FV space may be estimated by the *degree of heterogeneity* of the *distances between cluster centers* calculated over the given FV space. The intuition behind the hypothesis is that a FV space of high discrimination power shows a uniform (unbiased) distribution of distances between cluster centers, where each distance is equally likely to occur. The hypothesis was inspired by a basic image processing technique where the *contrast* in a gray image is estimated by the uniformity of its gray value histogram. The estimator is based on a compressed representation of the feature space to provide for noise suppression, and to simplify distance histogram calculation.
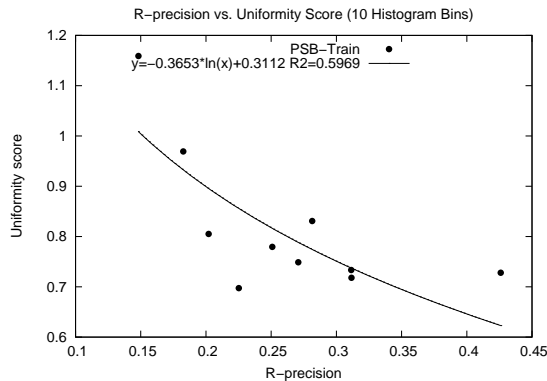

Figure 2: Regression analysis for the FV selection experiment described in Section 3.1.

An unsupervised estimator of discrimination power can be implemented as follows [9]. For clustering, we employ the *Self-Organizing Map* (SOM) algorithm [7] as a robust clustering algorithm that has been successfully adopted in many data analysis and visualization applications [11]. The SOM is a neural-network learning algorithm which for any

input set of data vectors calculates a fixed number of cluster prototypes arranged on a regular grid. Practically, we use the SOM algorithm to first cluster a candidate FV space f (a given data set in a candidate FV representation). We then calculate the histogram $h(f, b)$ of b bins length over the L_1 distances between all pairs of cluster prototypes adjacent on the SOM grid. Having obtained $h(f, b)$, we determine its corresponding degree of uniformity (uniformity score us) simply as its L_1 distance to the theoretic uniform histogram: $us(h) = \sum_{i=1}^{b} |h_i - \frac{1}{b}|$. Lower us scores indicate h to be more uniform-like, leading us to estimate higher discrimination power for the underlying FV space.

3. EVALUATION ON REAL DATA

3.1 Feature Extractor Selection Problem

We designed and conducted a series of experiments to systematically analyze the performance of the unsupervised discrimination estimator. Remember the goal is to mimic supervised benchmarking in FV design using automatic analysis. We therefore compare the results of the estimator against supervised benchmarking on a range of feature optimization problems.

We first test the estimator on a benchmark database of 3D models - the *Princeton Shape Benchmark* (PSB) Train partition [10] - indexed by a set of ten competing 3D FV extractors. These FV extractors are based on various geometric 3D features as described in [3] and elsewhere. As supervised reference benchmark score we use the R-Precision score [2] calculated over the benchmark classification information. R-Precision is a measure of retrieval precision when answering similarity queries under a given FV representation. Higher R-Precision scores indicate higher discrimination power of the underlying FV space. To assess the effectiveness of the unsupervised estimator, we compare it against respective R-Precision scores by correlation analysis. Figure 1 (a) plots the supervised R-precision scores of ten different FV spaces [4, 3]. Each method was benchmarked at several different resolution levels of the respective feature extractors, resulting in the shown R-Precision curves. Note there exists a FV-dependent optimal dimensionality setting for each FV.

We first evaluate the uniformity score in a *feature selec-*

(a) PMOM (15%, 1.16)   (b) SD2 (18%, 0.97)   (c) H3D (20%, 0.81)   (d) VOX (31%, 0.73)   (e) DBF (31%, 0.72)   (f) DSR (43%, 0.73)

(g) equal_1 (1.0, 1.80)   (h) equal_2 (2.0, 1.62)   (i) equal_3 (3.0, 1.28)   (j) equal_4 (4.0, 1.02)   (k) equal_5 (5.0, 0.84)
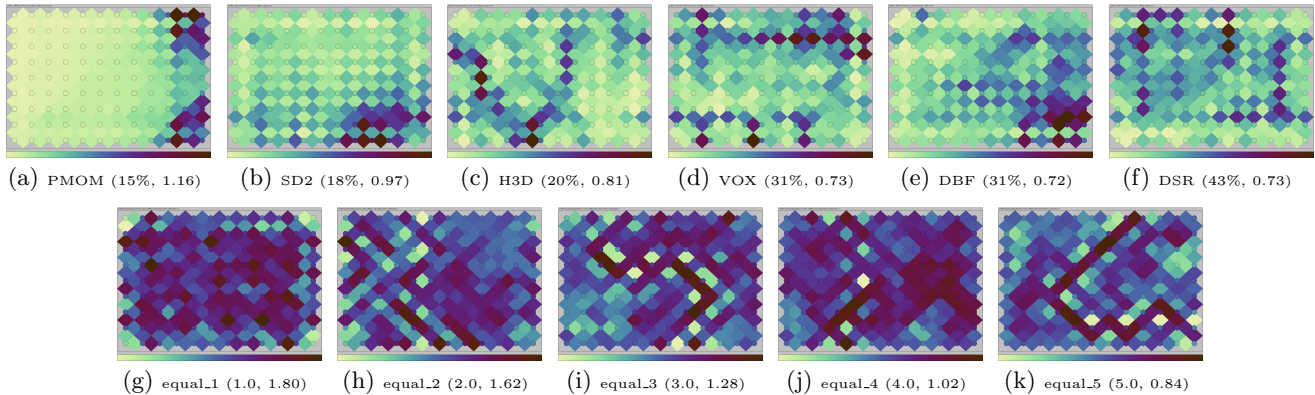
Figure 3: Top row: Inter-cluster distances for the FV spaces of lowest (a-c) and highest (d-f) discrimination power (R-precision and uniformity scores given in brackets). Bottom row: The same visualization for synthetic data sets under the *equal-sampling* scenario (d_span uniformity scores given in brackets).

tion experiment. We use the uniformity score estimator to determine the best discriminating FV representation. We consider the optimum dimensionality setting of each FV, according to Figure 1 (a). To evaluate the estimator, for each FV space, we calculated SOMs of dimensionality 12×9 for the PSB data set. Figure 2 plots the R-precision scores observed for the ten FV spaces against their respective uniformity scores (we set $b = 10$ to calculate the us score). The Figure also shows a logarithm model regression analysis for the plot. We verify the correlation between the supervised and the unsupervised FV space metrics at squared correlation coefficient $R^2 = 0.60$ (a measure for the quality of the dependence). While this is not a perfect functional dependency, both metrics clearly correlate in the expected sense: In FV spaces of better discrimination power (according to supervised benchmarking), we also observe more uniform distance distributions in the clustered FV space representations. The unsupervised us metric is therefore recommended to automatically select a discriminating FV representation for the database from a set of candidate FV spaces, *without* relying on supervised information: We select the FV space that minimizes the us score, thereby maximizing the distance distribution uniformity.

The analysis result can be illustrated by visualizing the underlying SOM-based distance distributions in form of so-called U-Matrices [11]. Figures 3 (a-f) show the distributions of inter-cluster distances for the three FV spaces with lowest and highest discrimination power, respectively. The images visualize the linearly (min,max) normalized magnitudes of distances between adjacent cluster prototypes in form of a diamond plot. It is obvious that the methods of lowest benchmarked discrimination power exhibit the most biased distance distributions. Specifically, image (a) is dominated by low distances, while the other images show increasingly more distance diversity, which is associated with higher expected discrimination power by the analysis function. Uniformity scores and R-precision scores are included in the figure for comparison. The visualization is useful for inspection of different FV spaces, and to understand the choice of the automatic feature selection scheme.

The unsupervised discrimination estimation may not only serve to select a single (best scored) discriminating feature vector to use. Using sets of feature vectors combining differ-

ent types of object characteristics often show superior results in practice, compared to using just a single type of features in a given application. The uniformity score may be used as a weight for combining multiple candidate feature vectors to an ensemble. E.g., the inverse of the uniformity score may serve as a weight in a linear combination of FVs. Feature vectors showing lower uniformity scores then enter the combination with higher weights, reflecting their better discrimination estimation. We state that we performed experiments using reciprocal us scores as weights for combining feature vectors, and observed an improvement in R-precision against the corresponding, unweighted ensembles.

3.2 Dimensionality Selection Problem

It is interesting to ask whether the estimator also works for the dimensionality selection problem, given a specific FV extractor, as these usually can be configured to different resolution levels, in itself being a feature selection problem traditionally solved by benchmarking. We conduct regression experiments by plotting the R-precision values obtained for different dimensionality settings of a given FV space against the uniformity scores of the respective settings. We selected a set of three FVs which allowed a sufficiently fine granularity of available dimensionality settings, and a significant spread in the R-Precision results for the sequence of dimensionality configurations. Figure 1 (b) shows the FV spaces and dimensionalities selected.

The charts in Figure 4 report the results of the experiment. The top row of charts shows plots of R-precision scores against uniformity scores, for all dimensionality settings of the given FV spaces. The uniformity scores were calculated using histogram width b such that the log regression dependency was maximized w.r.t. the R^2 statistic (values given in Figure 4). The dependency strengths range between 93% and 45%. The scatter plots indicate that a correlation exists between the supervised and the unsupervised FV metrics within each of these FV spaces for varying dimensionality settings. We observe that selecting the FV dimensionality minimizing the uniformity score not always manages to pick the optimal FV dimensionality (maximizing R-Precision), but gives solid selection results significantly outperforming the random choice.

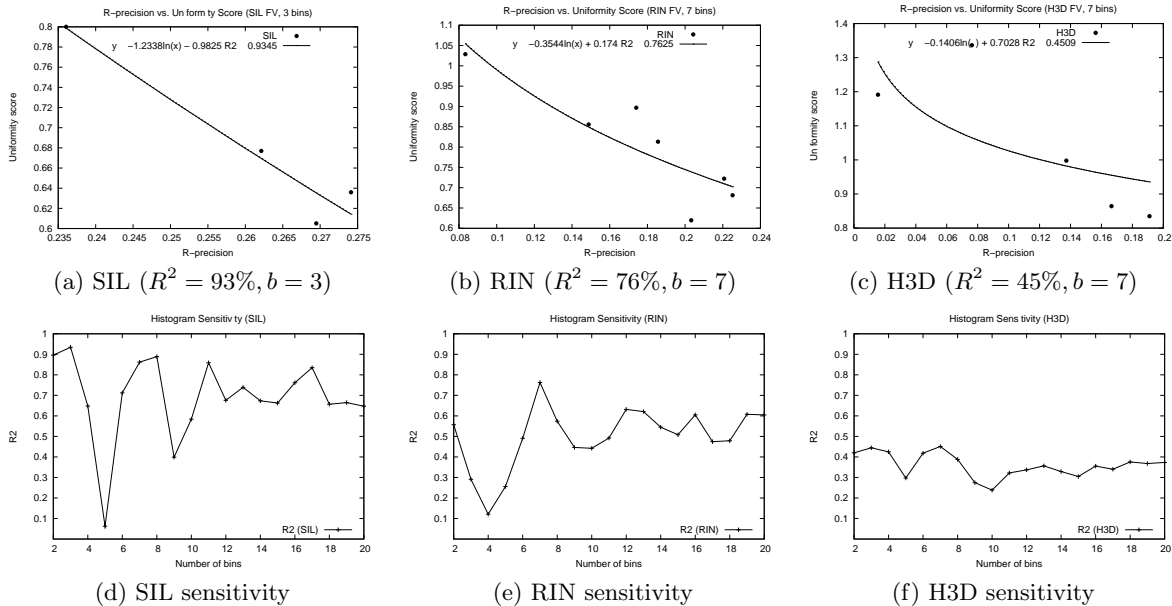A parameter of the method is the width of the histograms

| (a) SIL ($R^2 = 93\%, b = 3$) | (b) RIN ($R^2 = 76\%, b = 7$) | (c) H3D ($R^2 = 45\%, b = 7$) |
| (d) SIL sensitivity | (e) RIN sensitivity | (f) H3D sensitivity |

Figure 4: Results of the dimensionality selection experiment.

used to evaluate the uniformity score. We therefore investigated the robustness of the analysis with respect to bin width b, observing the correlation strength R^2 while varying b. The charts in the bottom row of Figure 4 plot the resulting R^2 values resulting for b in the interval $[2, 20]$. The correlation quality partially depends on the histogram width settings. The strongest dependencies result for small histogram widths between 3 and 7, depending on the FV space. For larger b settings, the correlation strengths stabilize at submaximal levels, showing certain oscillation. We currently attribute this observation to the simple equal-width binning approach used in our implementation of the uniformity score, possibly incurring discretization phenomena. We expect more advanced statistical uniformity tests to improve the analysis robustness.

4. EVALUATION ON SYNTHETIC DATA

The above findings link unsupervised information extracted from several given 3D FV spaces with respective discrimination power benchmarks. As benchmarks are supervised in nature and expensive to build, unsupervised discrimination estimation is desirable. We are therefore interested whether these results generalize. As the number of available multimedia benchmarks described by a significant number of different competing FV extractors is limited, we chose to generate synthetic data sets, simulating FV spaces of varying discrimination as follows:

- Dimensionality (dim) is set to 128, a medium sized resolution regarding many real FV extractors.

- Database size is set to 50 or 200 classes ($n\_classes$), each class consisting of 50 elements (c_size). These sizes are typical for real multimedia benchmarks.

- The FV dimensions are all normally distributed with $\sigma = 1.0$ around a class centroid.

- The class centers are uniformly distributed in FV space.

- Discrimination quality is controlled primarily by the size of the data space from which the class centroids are drawn. Specifically, we vary the span of all FV dimension (d_span) between 1.0 and 5.0.

We also control the size of the Self-Organizing Maps used for clustering prior to us score calculation. Specifically, we use SOM grid sizes of 12×9 and 32×24. Together with the above described parameters, we form three data scenarios by varying the proportion between the number of SOM prototype vectors allowed, and the number of classes contained in the synthetic benchmark. Table 1 gives the parameter settings for the scenarios. In the *equal-sampling* scenario, the number of SOM nodes is roughly equal to the number of classes (1.08 nodes/class). In the *over-sampling* and *under-sampling* scenarios, the number of SOM nodes is larger or smaller than the number of classes (15.36 and 0.54 nodes/class, respectively).

Table 1: Three synthetic data scenarios.

Scenario	dim	n_classes	c_size	d_span	SOM size
over	128	50	50	[1.0, 5.0]	32×24
equal	128	100	50	[1.0, 5.0]	12×9
under	128	200	50	[1.0, 5.0]	12×9

A series of experiments was run on these synthetic benchmark scenarios. Figures 5 (a-c) plot the uniformity scores (we set $b = 20$) obtained by varying d_span settings in the three scenarios. Consider plot (b) which gives the result for the equal-sampling scenario. There is a clear dependency between the two metrics: As class discrimination improves (d_span increases), the uniformity score us decreases. R^2 amounts to about 87%, indicating a strong correlation between class discrimination and uniformity score estimator. Figures (a) and (c) show the over-sampling and the under-sampling experiments. In both scenarios, we also observe a significant correlation between both metrics ($R^2 = 45\%$ and $R^2 = 70\%$, respectively). The strong correlation in the
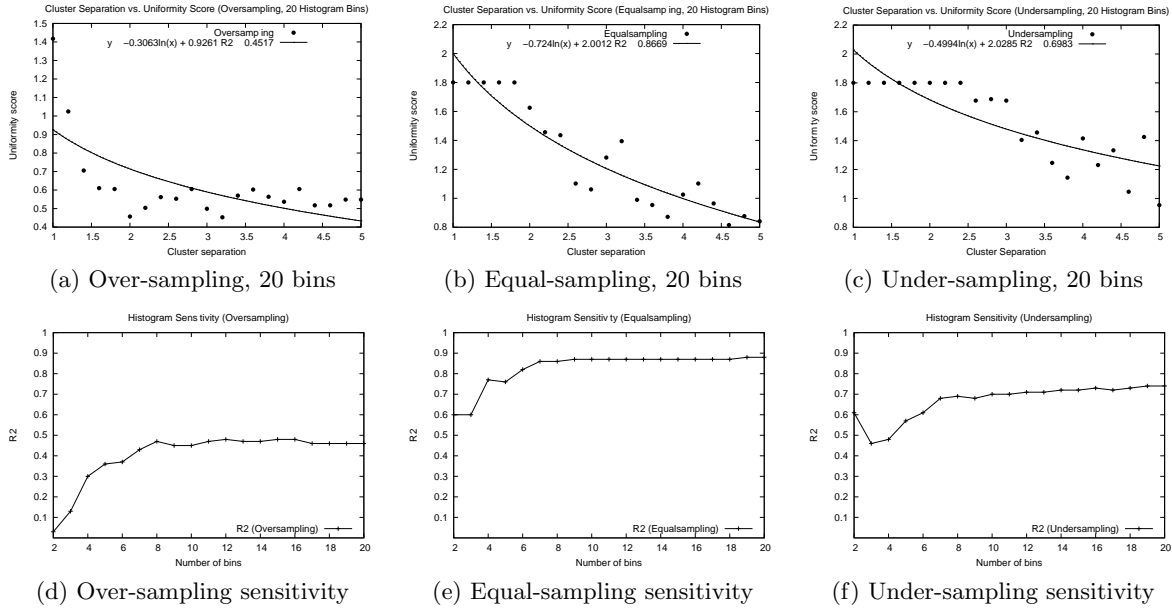
Figure 5: Results of the experiments on synthetic data.

under-sampling setting is specifically interesting to observe, as with 0.54 SOM prototype vectors per class, the number of SOM reference vectors is not sufficient to exclusively represent each class by a cluster center.

Again, we also varied the histogram length parameter b, to assess the robustness of the analysis. Figures 5 (d-f) report observed R^2 for the three scenarios when varying b between $[2, 20]$. For small sizes up to about 5 bins, R^2 is rather low (but increasing). We presume the small histogram sizes perform too much aggregation, eliminating useful information regarding the distance distributions. For $b = 7$ and above, we observe stable dependencies between $R^2 = 50\%$ and $R^2 = 85\%$, depending on the scenario.

In summary, these results indicate the robustness of the uniformity score-based discrimination analysis w.r.t. main parameter settings. Figure 3 shows the visualization of distance distributions in the equal-sampling scenario, for different d_span settings. Higher settings correspond to more discriminating FV spaces, and are reflected by more uniform distributions of distances.

5. CONCLUSIONS

We evaluated an automatic, unsupervised method for estimation of discrimination power in competing FV spaces. Systematic experiments on real and synthetic benchmark data showed that the unsupervised estimator based on distributions of distances in compressed FV space is able to closely resemble supervised benchmarking analysis. The method was shown to be robust w.r.t. main parameter settings. It is proposed as a valuable tool supporting and complementing the data-dependent and expensive supervised benchmarking approach to FV selection and optimization in multimedia applications.

Future work involves refining the method by including more sophisticated statistical distribution tests, experimenting with additional estimator functions, and evaluating the approach on additional benchmark data.

6. REFERENCES

[1] C. Aggarwal. On the effects of dimensionality reduction on high dimensional similarity search. In *Proc. ACM Symposium on Principles of Database Systems*, 2001.

[2] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, 1999.

[3] B. Bustos, D. Keim, D. Saupe, T. Schreck, and D. Vranić. Feature-based similarity search in 3D object databases. *ACM Computing Surveys*, 37:345–387, 2005.

[4] B. Bustos, D. Keim, D. Saupe, T. Schreck, and D. Vranic. An experimental effectiveness comparison of methods for 3D similarity search. *Int. Journal on Digital Libraries*, 6(1):39–54, 2006.

[5] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley-Interscience, New York, 2nd edition, 2001.

[6] A. Hinneburg, C. Aggarwal, and D. Keim. What is the nearest neighbor in high dimensional spaces? In *Proc. Int. Conf. on Very Large Data Bases*, 2000.

[7] T. Kohonen. *Self-Organizing Maps*. Springer, Berlin, 3rd edition, 2001.

[8] H. Müller, S. Marchand-Maillet, and T. Pun. The truth about corel - evaluation in image retrieval. In *Proc. Int. Conf. on Image and Video Retrieval*, 2002.

[9] T. Schreck, D. Keim, and C. Panse. Visual feature space analysis for unsupervised effectiveness estimation and feature engineering. In *Proc. IEEE Int. Conf. on Multimedia and Expo*, 2006.

[10] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The princeton shape benchmark. In *Proc. IEEE Int. Conf. on Shape Modeling and Applications*, 2004.

[11] J. Vesanto. SOM-based data visualization methods. *Intelligent Data Analysis*, 3(2):111–126, 1999.