A 3D Model Retrieval System via Statistical-Based Feature Selection

Kuan-Hsien Liu[1], Tsung-Jung Liu[2], and Hsin-Hua Liu[3]

[1]Assistant Professor, Department of Electrical Engineering, Chinese Culture University, Taiwan

[2]Assistant Professor, Department of Electrical Engineering, National Chung Hsing University, Taiwan

[3] Department of Electrical Engineering, National Taiwan University, Taiwan

Abstract - In this paper, we present a new 3D model retrieval system based on the statistical selected feature learning approach. This approach can be integrated with any existing and potential 3D model retrieval algorithm which includes 3D model feature extraction, selection and distance computation. By applying the analysis of variance of each component among all feature vectors and removing those components with F-values smaller than a critical F-value, we can reduce the feature dimension and also keep features with higher discriminating ability. Moreover, the SVM classifier is learned based on the ANOVA-selected features, which are extracted from the training set. We conduct experiments using the McGill Articulated Shape Benchmark database [1] for 3D model classification and retrieval, and demonstrate a significant performance improvement in the precision-recall curves.

Keywords: 3D model matching and retrieval, Analysis of variance, feature selection, shape, SVM classification.

I. INTRODUCTION

The number of 3D models has increased rapidly in the last decade. A large amount of research has been conducted on the development of an automatic 3D model retrieval system with a focus on retrieval accuracy characterized by precision and recall.

One of the key ingredients in a 3D model retrieval system is feature extraction. To be qualified as a good shape feature, it should possess high discriminant power and be invariant to various transformations. Generally speaking, shape features of 3D models can be categorized into several types; namely, statistical-based [2], visual-similarity-based [3][4], transform-based [5], and skeleton-based [6] methods.

The retrieval system computes the distance between any pair of shape features. If a pair of models is similar, the feature distance will be smaller, too. Hence, for a given 3D model, we can retrieve similar 3D models by computing the distance of its features and those of 3D models in the database. This is known as "content-based 3D model retrieval". We refer to [7]-[9] for detailed survey on this research topic.

There has been little work on feature dimension reduction in the context of content-based 3D model retrieval. One reason could be that the number of 3D models is large and it would be difficult to select a subset of features that are much more important than others a priori. In this work, we would like to address this problem from a new angel. That is, for a given set of models, we adopt a machine learning approach to learn the classifier in the training stage. Then, we use the obtained classifier to retrieve models in the test stage. There are two major advantages with the proposed approach: 1) lower complexity in the test stage and 2) better retrieval performance in terms of the precision-recall tradeoff.

The rest of this paper is organized as follows. We first give an overview of our proposed 3D model retrieval framework in Section II. Then, we review several well-known shape features for 3D model retrieval in Section III. The proposed class-dependent feature learning and statistical-based feature selection approach is described in Section IV. Experimental results are reported and performance evaluation is conducted in Section V. Finally, concluding remarks and future research topics are given in Section VI.

II. OVERVIEW OF PROPOSED 3D MODEL

In this section we give an overview on our proposed 3D model by briefly describing the function of each block of the framework shown in Figure 1. First, the 3D model features are extracted. The dimensions of the extracted features are high and we apply the statistical analysis to select some of the features to reduce the feature dimensionality, which will lower the complexity of the system. Then, the selected features are used for training a 3D model retrieval classifier and it will be used for testing the queried 3D models. Finally, the system output will give the retrieved 3D model.

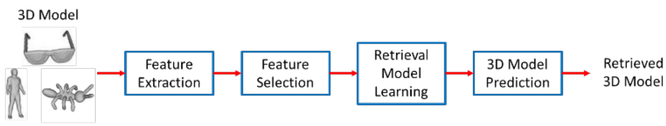


Fig. 1 Overview of proposed 3D model retrieval framework.

III. RELATED WORK

We review some of related work with a focus on the feature extraction and their distance computation used in 3D model matching and retrieval. A detailed comparison of 3D model retrieval methods for non-rigid 3D objects can be found [10]. Some very recent work: [11] uses Randomized sub-volume partitioning approaches and [12] uses Bag-of-View-Words for 3D model/object retrieval. Following is a review of six different feature extraction methods on 3D model retrieval.

AAD (Absolute Angle Distance) [13]. It is a method that computes the features by first converting a surface-based input model into an oriented point-set model and then computing joint 2D histogram of distance and orientation of pairs of points. The length of an AAD feature vector is 256.

D2 (Distance between 2 random points) [2]. It is a method that measures the histogram of Euclidean distances between pairs of randomly selected points on the surface of a 3D model. The number of histogram bins is chosen as 1024 so that the length of a D2 feature vector is 1024.

LFD (Light Field Descriptor) [3]. The light field cameras are put on 20 different views uniformly distributed over a 3D model. Since the silhouettes projected from two opposite vertices are identical, 10 different silhouettes are produced for a 3D model. To be robust against rotations among 3D models, a set of 10 LFDs is applied to each 3D model. Therefore, it is a method that represents a 3D model by 100 silhouettes (10 views per group) rendered from uniformly distributed viewpoints over a hemisphere and the silhouette is encoded by a feature vector with 47 entries including 35 Zernike moments, 10 Fourier coefficients, 1 eccentricity and 1 compactness. The length of an LFD feature vector is 4700. For any 3D model, even a simple one, 10 descriptors are created, and 10 silhouettes are represented for 20 viewpoints in each descriptor. Therefore, a total of 100 silhouettes will be rendered and the length of an LFD feature vector for any 3D model is 4700.

SPRH (Surflet-Pair-Relation Histograms) [14]. It uses the modified SPRH [15] to extract features of a 3D model. The length of a modified SPRH (**mSPRH**) feature vector is 625.

PS (linearly Parameterized Statistics) [16]. It is a method that uses a combination of three vectors (i.e., the moment of inertia, the average distance of surfaces from the axis, and the variance of distances of surfaces from the axis.) Values in each vector are discretely parameterized along each of the three principal axes of inertia of the 3D model. The length of a PS feature vector is 567.

SHD (Spherical Harmonic Descriptor) [5]. It is a method that describes a 3D model as a feature vector consisting of spherical harmonic coefficients, which are extracted from three spherical functions giving the maximal distance from the center of mass as a function of a spherical angle. The length of a SHD feature vector is 544.

The dimensions of feature vectors discussed in above are summarized in Table 1. After obtaining shape features of 3D models, we will analyze these features and select the subset of most discriminant features for a given class of models automatically using a machine learning approach as described in the following section.

TABLE 1. FEATURE VECTOR DIMENSIONS

Feature	AAD	D2	LFD	mSPRH	PS	SHD
Length (L)	256	1024	4700	625	567	544

IV. 3D MODEL FRAMEWORK

In this section we will describe the details of the proposed framework. The block-diagram of the proposed 3D model retrieval system is shown in Figure 2.

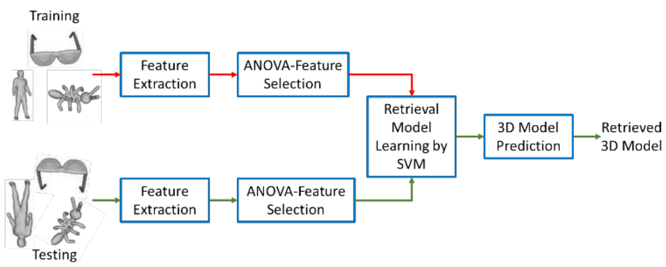


Fig. 2 Our proposed 3D model retrieval framework.

This framework consists of the following three main modules:

1. Pre-processing for feature dimension reduction, which is completed by ANOVA-feature selection [17];
2. Feature learning via support vector machine (SVM);
3. Post-processing of learned features.

They will be detailed in the following sub-sections.

A. *Feature reduction via selection*

The purpose of feature selection is to keep those features having higher discriminating power and discard those features having lower discriminating power. The dimension (i.e., the number of elements) of a feature vector may be very high since it depends on the type of features used to extract from 3D models. To reduce the dimension, we need to determine which features in a feature vector are significantly different across type groups of 3D models. Based on the idea of hypothesis testing, the "unpaired t test" [17] or the "analysis of variance (ANOVA)" [17] method can be used for the separation of two groups. However, to separate three or more than three groups, "ANOVA" is more suitable and adopted here. The procedure [18, 19] to identify which feature has higher discriminating power among groups using ANOVA is described below.

Given *m* groups and *n* 3D models per group, for a feature *X* in the feature vector, we calculate the following quantities:

- Mean of each group $\overline{x}_1, \overline{x}_2, \cdots, \overline{x}_m$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

- Variance of each group $s_1^2, s_2^2, \cdots, s_m^2$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{2}$$

- Within group variance

$$s_{within}^2 = \frac{1}{m}\sum_{i=1}^{m} s_i^2 \tag{3}$$

- Overall mean

$$\overline{X} = \frac{1}{m}\sum_{i=1}^{m} \overline{x}_i \tag{4}$$

- Standard error of the mean

$$s_{\overline{X}}^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\overline{x}_i - \overline{X}\right)^2 \tag{5}$$

- Between groups variance

$$s_{between}^2 = n\, s_{\overline{X}}^2 \tag{6}$$

- *F* statistic value

$$F = \frac{s_{between}^2}{s_{within}^2} \tag{7}$$

- Degree of freedom

$$v_n = m - 1$$

$$v_d = m(n-1) \tag{8}$$

- *F* critical

$$F_{crit} \equiv F_\alpha(v_n, v_d) \tag{9}$$

where $F_\alpha(v_n, v_d)$ can be obtained from Table 3-1 on [17].

If $F > F_{crit}$, we reject the null hypothesis H_0: {There is no significant difference on feature X between different groups} with $P < \alpha$, and α is the significance level, which is usually set to 0.05 or 0.01. Hence, we select those features with higher *F* values to get the selected feature vector.

B. *SVM classification*

In this module, we would like to explain how to get the SVM classifier from the training data. The training, testing, and cross-validation steps are described as follows.

Feature Vector Labeling. Label each feature vector, which is a row of the feature matrix, with value *i* if the model belongs to class *i*.

Linear Scaling. Linearly scale training and testing data. Every entry in a feature vector is a sub-feature. We scale each column linearly to range [0, 1]. This is conducted to avoid the dominance of attributes with a large dynamic range over those with a smaller dynamic range.

N-Fold Cross-Validation. We divide the entire database of 3D models N subsets of equal size (or nearly equal size) where each subset consists of about the same number of 3D models from each class. Then, we choose 1 subset as the testing set while using the other N-1 subsets as the training set. This process is repeated for N times where each subset is used as the testing set once. The technique, called the N-fold cross-validation, is employed to average the testing results and increase the confidence level.

Kernel Selection. Given a training set of feature-label pairs $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, l$, where $\mathbf{x}_i \in R^n$ and $y_i \in \{1, -1\}^l$, in the SVM, the training feature vectors $\mathbf{x}_i$ are mapped into a higher dimensional space by function φ. Furthermore, $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$ is called the kernel function. Two commonly used kernel functions are

- Linear:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i{}^T \mathbf{x}_j \tag{10}$$

- Radial basis function:

$$K(\mathbf{x}_i, \mathbf{x}_j) = exp\left(-\gamma\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right), \gamma > 0 \tag{11}$$

where γ is the kernel parameter.

The radial basis function (RBF) kernel is often used as the kernel when the dimension of the feature vector is low [20, 21, 22]. On the other hand, if the dimension of the feature vector is high, which is our current case, the nonlinear mapping does not improve the performance much. Thus, we choose the linear kernel for the SVM algorithm in our experiment.

C. *3D Model Classification*

Every 3D model needs an index value to represent itself. Here, we assign a 3D model with the following new index:

$$F = i^{*}\cdot 10 + r, \tag{13}$$

where i^* is the new class index number and r is a random number in the unit interval (0,1). The reason to multiply the new class index i^* by 10 is to separate the index value of 3D models in different classes. We can view F as one additional feature of the K-SVM method, where K could be any of D2, LFD, SHD, PS, AAD and mSPRH in our experiments.

When each 3D model is predicted with a new class index i^*, the classification accuracy can be determined by comparing the predicted new class index and its ground-truth. If $i = i^*$, it means that this is a correct classification result. Otherwise, it is a wrong classification result. The distance matrix can then be constructed by calculating the distance between every pair of 3D models' new feature F. Thus, we can plot performance curves such as the precision-recall (P-R) curves accordingly.

V. Experimental Results

In this section, we will demonstrate our experiment, including database, experiment setup, classification accuracy, and retrieval performance evaluation.

A. *Database and Experiment Setup*

We choose the McGill Articulated 3D model database, which contains 255 models with 10 classes in our experiments. The feature matrix has a size of 255 x L, where L is the length of the feature vector. In the training of SVM classifiers, we use the LIBSVM library [23]. In the cross-validation step, we divide the entire set of 3D models into N = 5 subsets. One subset is sequentially tested using the classifier trained based on the remaining 4 subsets. We will report the 3D model classification and retrieval performance in Sec. V.A and Sec. V.B, respectively.

B. *Classification Performance*

First, we study the performance of 3D model classification and compare the accuracy and complexity. We implement the 6 methods as described in Section III. Furthermore, we implement the proposed dimension-reduced features as well as the SVM-based feature training and testing process in association with each method [24]. In the feature dimension reduction process, we discard columns having the largest variance values gradually and find the best classification accuracy with respect to the number of columns. Thus, it is a result obtained from exhaustive search. The results are showed in Table 2.

As shown in Table 2, we see that the use of the dimension-reduced feature to learn a 3D model classifier improves the classification accuracy as well as reduces the training and testing complexity.

TABLE 2. ACCURACY AND COMPLEXITY OF 3D MODEL RETRIEVAL

Feature type	**Original feature**		**ANOVA-selected feature**		
	Dim.	**Accuracy**	**Dim.**	**Accuracy**	**Reduced feature dimension**
AAD	256	92.94% (237/255)	186	94.11% (240/255)	27 %
D2	1024	79.60% (203/255)	658	84.31% (215/255)	36 %
LFD	4700	89.01% (227/255)	2980	92.15% (235/255)	37 %
mSPRH	625	7.84% (20/255)	451	94.50% (241/255)	28 %
PS	567	62.35% (159/255)	393	70.19% (179/255)	31%
SHD	544	89.41% (228/255)	376	92.94% (237/255)	31 %

C. *Retrieval Performance*

Next, we examine the performance of content-based 3D model retrieval. The precision-recall plot is a common tool in evaluating the retrieval performance. For each query model in class i and any number N of top matches, "recall" and "precision" are defined as [7]:

$$recall = \frac{models\ in\ class\ i\ returned\ within\ the\ top\ N\ matches}{number\ of\ models\ in\ class\ i}$$

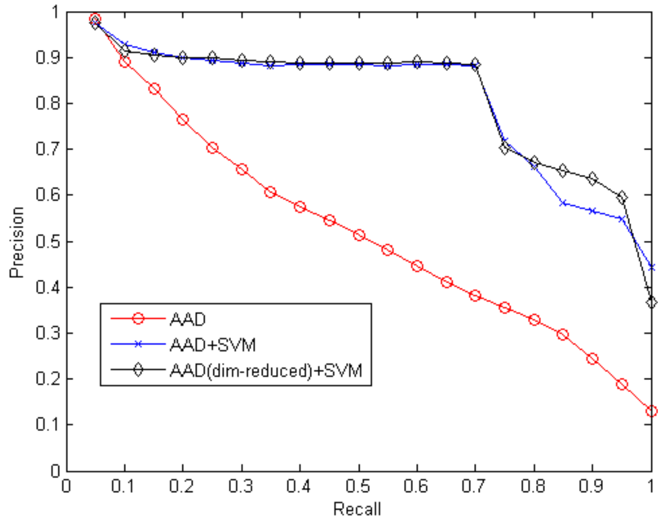
$$precision = \frac{the\ top\ N\ matches\ that\ are\ members\ of\ class\ i}{the\ top\ N\ matches}$$

A perfect retrieval result will give a horizontal line across the top of the plot (with precision = 1). Thus, a curve that lies more towards the upper right position indicates a better retrieval performance. We compare the retrieval performance in terms of the precision-recall curves for each of the 6 methods described in Section III with three variants:

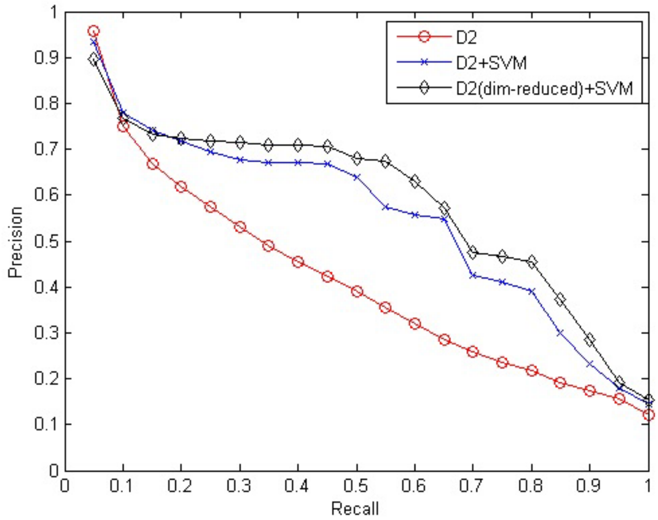
i. the original method (K);
ii. its improved version by incorporating SVM (K+SVM);
iii. its improved version by incorporating feature dimension reduction and SVM (K+FDR+SVM).

The results are shown in Figure 3, where the 6 original methods are AAD, D2, LFD, mSPRH, PS, and SHD (from (a) to (f)).

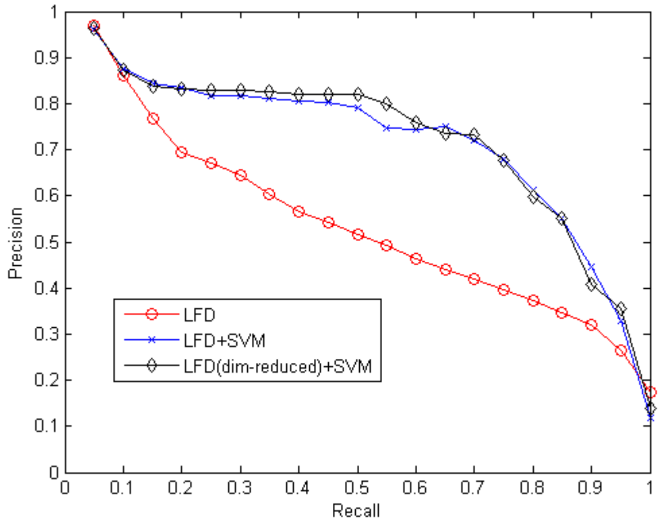
As compared with the original method, we observe a remarkable improvement in the retrieval performance by incorporating SVM and/or joint FDR/SVM. It is also interesting to point out that, although the use of either SVM alone or joint FDR/SVM offers similar improvement for most methods, it demands joint FDR/SVM to achieve performance improvement for mSPRH.



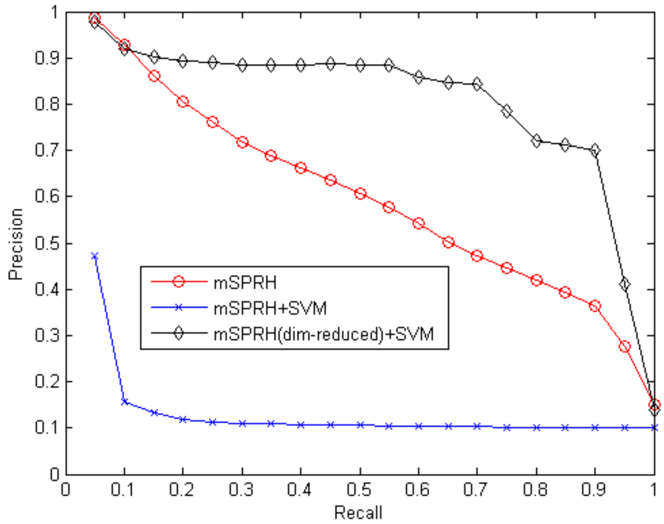
(a)



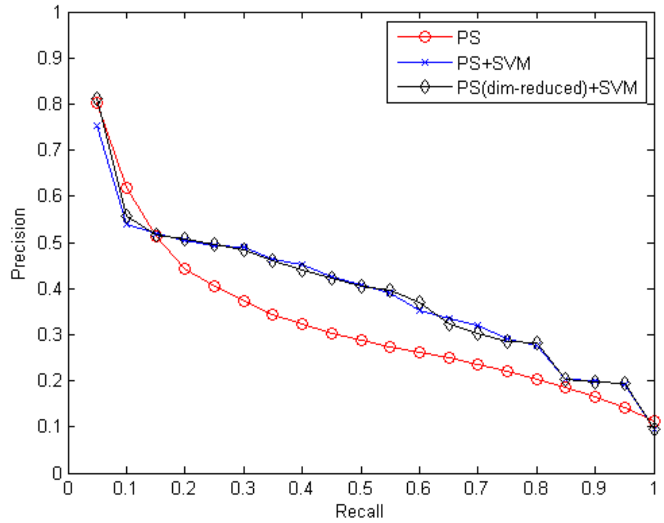
(b)



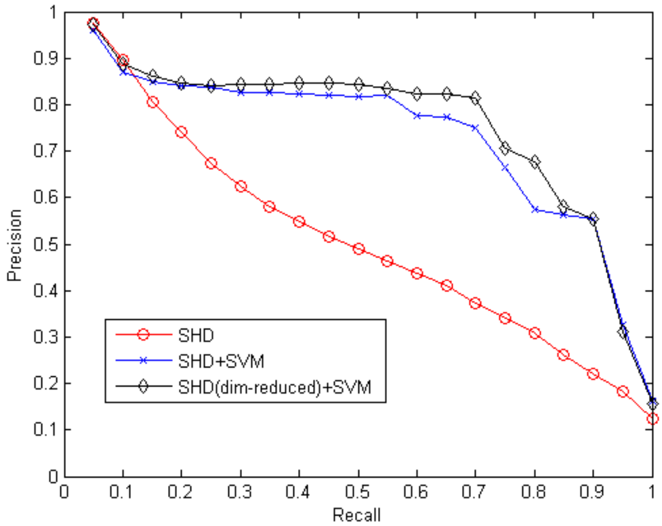
(c)



(d)



(e)



(f)

Fig. 3 Comparison of the precision-recall curves for three methods (original feature sets, original feature sets with SVM training, and original feature set with feature dimension reduction and SVM training.

Since some of the mSPRH features for 255 models from different classes are similar, it affects the training of the

SVM classifier. This could be the reason why the performance of mSPRH+SVM is much worse than that of mSPRH. With the FDR in mSPRH, the similarities between those similar features will be greatly reduced, and hence the mSPRH+FDR+SVM performs much better than mSPRH.

VI. CONCLUSION

In this work, we propose an SVM-based 3D model retrieval system with statistical feature selection. Our system can analyze the feature set provided by any other methods and reduce its dimension based on the idea of increasing the discriminating information. The SVM algorithm is used to train a classifier and a cross-validation technique is employed to increase prediction reliability. Experimental results were given to demonstrate the superior performance of the proposed approach. Since the classification technique is built upon a well-trained classifier, it can possess much better discrimination ability, which is verified in the precision-recall plots. This work also shows any 3D model retrieval methods can adopt our framework to enhance their performance.

REFERENCES

[1] K. Siddiqi, J. Zhang, D. Maxrini, A. Shokoufandeh, S. Bouix, S. Dickinson, "Retrieving articulated 3D models using medial surfaces," in *Machine Vision and Applications*, vol. 19, no. 4, pp. 261–274, 2008.

[2] R. Osada, T. Funkhouser, B Chazelle, and D Dobkin, "Shape distributions," in *ACM TOG*, vol. 21, no. 4, pp. 807–832, 2002.

[3] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, "On visual similarity based 3D model retrieval," in *Proc. Eurographics 2003*, vol. 22, pp. 223–232, 2003.

[4] Z. Lian, A. Godil, and X. Sun, "Visual Similarity based 3D Shape Retrieval Using Bag-of-Features," in *Proc. SMI'10*, pp. 25-36, 2010.

[5] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Proc. SGP'03*, vol. 43, pp 156–164, 2003.

[6] H. Sundar, D. Silver, N. Gavani, and S. Dickinson, "Skeleton based shape matching and retrieval," in *Proc. SMI'03*, pp. 130–139, 2003.

[7] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton Shape Benchmark," in *Proc. SMI'04*, pp. 167–178, 2004.

[8] Y. Yang, H. Lin, and Y. Zhang, "Content-based 3-D model retrieval: A survey," in *IEEE Transactions on Systems, Man, and Cybernetics - Part C*: Applications and Reviews, vol. 37, no. 6, Nov. 2007.

[9] J. W. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," in *Multimedia Tools and Applications*, vol. 39, no. 3, pp. 441–471, 2008.

[10] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoué, H. Van Nguyen, R. Ohbuchi, and Y. Ohkita, "A comparison of methods for non-rigid 3D shape retrieval." in *Pattern Recognition*, vol. 46, no. 1, pp. 449-461, 2013.

[11] T. Furuya, K. Seiya, and O. Ryutarou. "Randomized sub-volume partitioning for part-based 3D model retrieval." in *Proceedings of the 2015 Eurographics Workshop on 3D Object Retrieval*, pp. 15-22. Eurographics Association, 2015.

[12] K. Ding, W. Wang, and Y. Liu. "3D model retrieval using Bag-of-View-Words." *Multimedia Tools and Applications, vol. 72, no. 3, pp. 2701-2722*, 2014.

[13] R. Ohbuchi, T. Minamitani, and T. Takei, "Shape-similarity search of 3D models by using enhanced shape functions," *International Journal of Computer Applications in Technology (IJCAT)*, Vol.23, No. 2/3/4, pp. 70-85, 2005.

[14] E. Wahl, U. Hillenbrand, and G. Hirzinger, "Surflet-Pair-Relation Histograms: A Statistical 3D-Shape Representation for Rapid Classification," in *proceedings Forth International Conference on 3-D Digital Imaging and Modeling (3DIM 2003)*, pp. 474-481, 2003.

[15] R. Ohbuchi and Y. Hata, "Combining Multiresolution Shape Descriptors for 3D Model Retrieval," in *Proc. WSCG'2006*, 2006.

[16] R. Ohbuchi, T. Otagiri, M. Ibato, and T. Takei, "Shape-similarity search of three-dimensional models using parameterized statistics," in the *proceedings of the Pacific Graphics*, 2002.

[17] S. A. Glantz, Primer of Biostatistics. McGraw Hill, 6th edition, 2005.

[18] K.-H. Liu, S. Yan, and C.-C. J. Kuo, "Age group classification via structured fusion of uncertainty-driven shape features and selected surface features," in *IEEE Conference on. Winter Applications of Computer Vision (WACV)*, pp. 445 – 452, 2014.

[19] K.-H. Liu, S. Yan, and C.-C. J. Kuo, "Age estimation via grouping and decision fusion," in *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2408–2423, 2015.

[20] K.-H. Liu, T.-J. Liu, and H.-H. Liu, "A sift descriptor based method for global disparity vector estimation in multiview video coding," in *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1214–1218, 2010.

[21] T.-J. Liu, K.-H. Liu, J. Y. Lin, W. Lin, and C.-C. J. Kuo, "A paraboost method to image quality assessment," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. x, no. x, pp. xx – xx, 2016. (to appear)

[22] K.-H. Liu, T.-J. Liu, H.-H. Liu, and S.-C. Pei, "Facial makeup detection via selected gradient orientation of entropy information," in *IEEE International Conference on Image Processing (ICIP)*, pp. 4067–4071, 2015.

[23] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," in *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, 27, 2011.

[24] T.-J. Liu, K.-H. Liu, H.-H. Liu, and S.-C. Pei, "Age estimation via fusion of multiple binary age grouping systems," in *IEEE International Conference on Image Processing (ICIP)*, pp. 609–613, 2016.