FEATURE MATCHING OF MULTI-VIEW 3D MODELS BASED ON HASH BINARY ENCODING

H. Li[*], T. Zhao[*], N. Li[†], Q. Cai[*], J. Du[‡]

Abstract: Image data and 3D model data have emerged as resourceful foundation for information with proliferation of image capturing devices and social media. In this paper, a feature matching method based on hash binary encoding for multi-view 3D models in social media is proposed. SIFT algorithm is first used to extract features of the depth image, and then RANSAC is utilized as a filter. Finally, a cascade hash binary encoding algorithm is adapted to match the feature of multi-view models. Experimental results on SHREC2014 dataset have shown the effectiveness of the proposed method.

1. Introduction

Online social networking is attracting more and more attention in today's Internet, where users can share and consume all kinds of multimedia contents. The contents range from 3D models, images, and videos, to text available on the Web or kept in storage as a collection of big data [2]. Information retrieval has been developed from a single text object to 2D image or other multimedia information, and even more realistic 3D models or 3D scene. Furthermore, 3D model has been widely used in many fields such as CAD, 3D games, medical image, art restoration and so on [12, 14, 27]. However, it is a challenging task to gain a desired model in a mass retrieval database due to the explosive growth of the 3D models in recent years.

There are many kinds of classification for the 3D model retrieval [22]. Based on the types of features employed, existing 3D model retrieval techniques can be classified into four categories: geometry-based, graph-based, view-based, and hybrid

[*]Haisheng Li – Corresponding author; Tianyu Zhao; Qiang Cai; Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, No. 11, Fucheng Road, Haidian District, Beijing, China, E-mail: lihsh@th.btbu.edu.cn

[†]Nan Li; School of Materials Science and Mechanical Engineering, Beijing Technology and Business University, No. 11, Fucheng Road, Haidian District, Beijing, China,

[‡]Junping Du; School of Computer Science and Technology, Beijing University of Posts and Telecommunications, No. 10 Xitucheng Road, Haidian District, Beijing, China,

techniques. In the view-based methods, 3D objects are represented by a group of images from different directions. These images are captured with a static camera or virtual camera array. As each object is represented by a set of multiple views instead of original 3D model, some existing image processing/matching methods can be employed. In view-based techniques, a 3D model is represented by a set of rendered views. The visual similarity between the views of two models is regarded as the model difference. An example of multi-view 3D model is shown in Fig. 1.
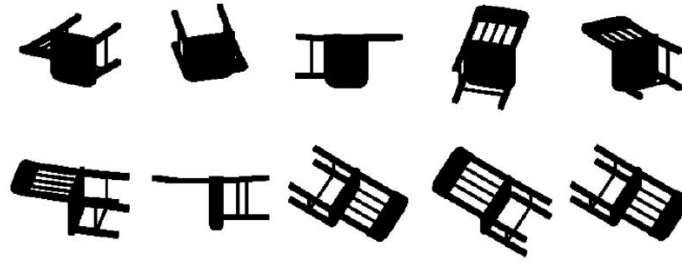

Fig. 1 *An example of multi-view 3D model.*

2. Related works

In this paper, general RGB images without background and the projection rendering silhouette images, contour images and depth images of 3D model are referred as views.

Generally, the extracted features from the depth images are as follows: depth buffer descriptor [25], the two-dimensional discrete wavelet transform [21], aligned multi-view depth line feature descriptor [3], view-based PCA feature descriptor [24]. DSIFT feature descriptor [18,23] and 3D surf descriptor [10,11] etc.

Lindstrom et al. [13] presented notion of image-driven simplification, a framework that uses images to decide which portions of a model to simplify. Mahmoudi et al. [17] applied two value images of 3D model projection apply to 3D model shape retrieval. Chen et al. [4] proposed a Light Field Descriptor (LFD). Vrani C et al. [25] suggested to regard 3D model projection as a set of depth images (6 images), and put forward a depth image of the Depth buffer-based descriptor. Ansary et al. [1] proposed an adaptive views clustering (AVC) algorithm which could achieve 3D model retrieval. Papadakis et al. [21] proposed an unsupervised retrieval algorithm based on panoramic view. Gao et al. [8] proposed a SSCD (spatial structure circular descriptor) descriptor which 3D model are projected to its minimum outer bounding sphere. Ohbuchi et al. [19] adapted the SIFT descriptor to depth image and proposed an approach using DSIFT, SSIFT, and 1-SIFT to achieve good results. Similarly, Wang et al. [26] proposed a discriminative probabilistic object modeling and defined the upper bound of the Kullback-Leibler divergence as the distance of two object to accomplish the retrieval.

3.   The proposed method

3.1   Feature extraction

We first perform image normalization to eliminate the influence of linear and non-linear illumination on the images as a preprocessing. We have a set of depth images whose foreground have already been extracted. Laplacian operator is used to sharpen depth image, then SIFT algorithm is adopted to get the key points. An algorithm called RANSAC [6] (Random Sample Consensus) is employed as the filter before feature matching. It can efficiently remove non-related matching points, and thus improve the matching accuracy. The processing of feature extraction and feature matching is shown in Fig. 2.
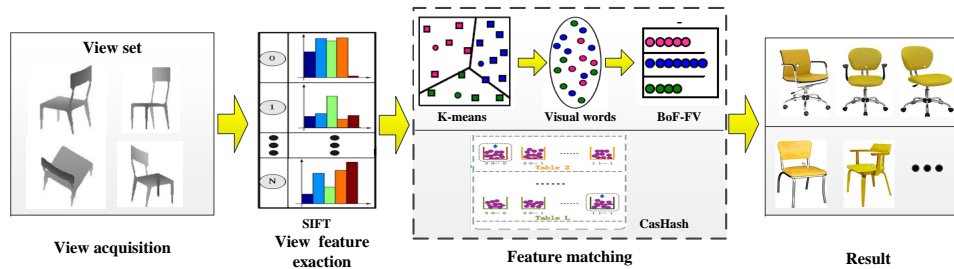

Fig. 2 *The process of the proposed algorithm.*

3.1.1   Image sharpening

During the process of image sharpening, the edge of the image gray transition can be sharpened. Laplacian operator is one of isotropic derivative operator in second differential equations. We use the Laplacian operator for image sharpening.

3.1.2   SIFT algorithm

SIFT (Scale Invariant Feature Transform) algorithm was developed by David Lowe [15]. The SIFT descriptor is robustness against partial occlusion, clutter, noise, lighting changes, and geometric translations. This algorithm is widely used in computer vision related to point matching between different views of a 3D scene and view-based object recognition [16]. We used SIFT to create the various feature vectors for each image.

3.1.3   Sample consensus algorithm

RANSAC (Random Sample Consensus) algorithm was proposed by Fischler and Bolles [6], also referred as a random sample of the consistency algorithm, which is a robust parameter estimation method. A set of observed data sets are mainly utilized to fit a mathematical model on the iterative method, followed by random sampling to remove the uncertainty points in the image.

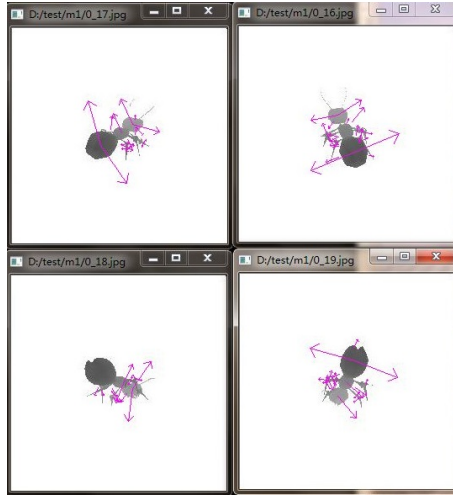The result of feature extraction by using SIFT algorithm is shown in Fig. 3.

Fig. 3 *SIFT algorithm result.*

3.2 Feature matching

A cascade matching hash binary algorithm (CasHash algorithm) is adapted for the feature matching process. Fig. 4 shows the processing of the feature matching.
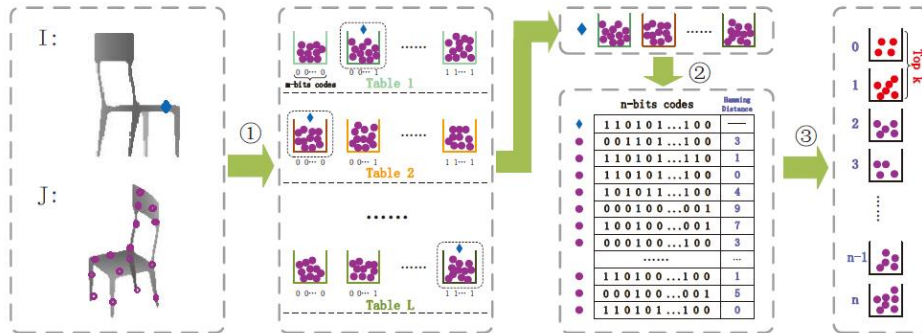

Fig. 4 *The process of Multi-View 3D model feature matching.*

The locality-sensitive hashing (LSH) algorithm is another approximating nearest neighbor search algorithm based on hash index. LSH can reduce the dimensionality of high-dimensional data. LSH hashes map similar items to the same "bucket". As shown in Fig. 5, the locality-sensitive hashing algorithm is designed as Eq. (1):

$$h_r(\mathbf{q}) = \begin{cases} 1 & \mathbf{r} \cdot \mathbf{q} > 0 \\ 0 & \mathbf{r} \cdot \mathbf{q} < 0 \end{cases},$$

where $\mathbf{r}$ is a d dimensional random vector which obeys the normal distribution $N(0,1)$. And $\mathbf{q}$ is a depth image feature vector to be binarized.
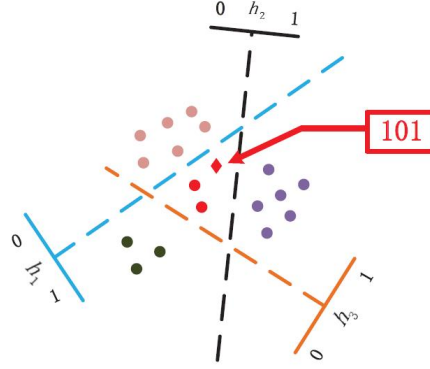
Fig. 5 *Locality-sensitive hashing (LSH).*

Distance formula is shown as Eq. (2):

$$D(x, y) = \frac{\theta(x, y)}{\pi}, \tag{2}$$

where $\theta(x, y)$ is the cosine distance.

The proposed Cascade Hashing structure contains three layers which map the 3D model representation into binary codes from coarse to fine, resulting in significant speedup of model matching. In addition, each layer of the Cascade Hashing adopts different measures and filtering strategies, which is demonstrated to be less sensitive to the noise of feature points.

CasHash algorithm consists of three steps. They are Hashing lookup, Hashing remapping and Hashing ranking. The 0–1 binary coding granularity of these three hash maps are from coarse to fine, that is to say, feature dimension is increasing.

The essence of CasHash algorithm is that for each feature point on the image I (the feature vectors of the SIFT key points), three steps of the CasHash were applied to return the matching points of the image J. The process of finding the matching point in image J for the feature point p in image I is as follows.

(1) Hashing lookup: LSH is used to make the feature point p and all the feature points in the image J for the low-dimensional binary coding. Assumeit is m-dimensional. The feature points with same coding were put into the same bucket, the feature point p and the matching points which fall into the same bucket on the image J return as candidate feature points. We proposed the low-dimensional and multi-table method to increase the number of the feature points which fall into the same bucket. Supposed the number of table is L, an $m \times L$ d-dimensional independent random vectors $\mathbf{r}$ is needed to be built. This step is a coarse-search process. Experiments have shown that when m was taken 8 and $\mathbf{r}$ was taken 6, a better result could be obtained.

(2) Hashing remapping: Due to the huge number of candidate feature points, LSH is used to remap the feature point and candidate feature point set into

hamming space which is high-dimensional coding. The proposed dimensional of the remap is 128.

(3) Top K ranking: We build the hash table by using the hamming distance of the feature point p and candidate feature point set. Firstly, the distance is sorted as the key value. The feature points with the same key value will fall into the same hash bucket. Then, we traverse the hash table from front to back, until we get the K feature points. Finally, we propose the Euclidean distance to get the Top 2 in the K candidate set. If these two feature points pass the Lowes ratio test, the matching point of the feature point p on the image J is found.

In this paper, we transformed the number of the key points in the two depth images into the similarity distance (hausdorff distance) between the two depth images. And based on the transformation, the matching method between view sets are given. The time complexity of CasHash algorithm is shown as Eq. (3):

$$O\left(T_H \cdot LN^2/2^m + T_E \cdot Nk + dmLN + ndLN/2^m\right). \tag{3}$$

In Hashing lookup step, the time complexity of hausdorff distance for $LN/2^m$ points is $O(T_H \cdot LN^2/2^m)$ and the comlexity of the hash is $O(dmLN)$. In Hashing remapping step, the time complexity of the hash is $O(ndLN/2^m)$. In the final Top K ranking step, it costs $O(T_E \cdot k)$ (Euclidean distance is represented by E).

4. Experiments

SHREC 2014 LSSTB TARGET MODELS [11] is chosen as the experimental dataset. This dataset contains 171 classes and 8987 models. Each class has 53 models. The number of vertices of each model is 5233. The model files are stored into OFF format which are ASCII text files. Fig. 6 shows 8 types of models in the dataset.
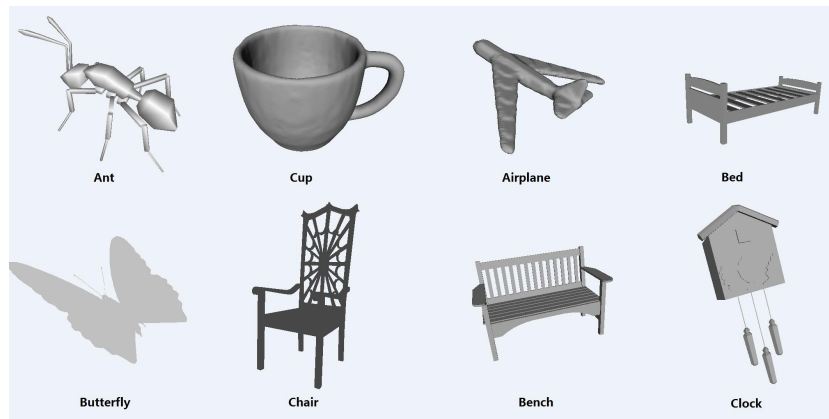

Fig. 6 *Example models in the dataset.*

We have adopted six evaluation criteria which have been widely used in the field of 3D shape retrieval: Precision-Recall, Nearest neighbor, First-tier, Second-tier, E-Measure, Discounted Cumulative Gain [11].

- Precision-Recall plot (PR): Assume there are n models in the dataset, precision P is to measure the accuracy of the relevant models among the top $K(1 \leq K \leq n)$ ranking results, while recall R is the percentage of the relevant class that has been retrieved in the top K results.

- Nearest neighbor: the rate of the closest models that belong to the same class as the inquiry. The maximum score is 100 %, a higher score shows a better result.

- First-tier, Second-tier: The rate of models in the inquiry's class that appear within the top K models and the K depends on the size of the inquiry's class. If a class get C members, the first tier is $K = C - 1$, and the second tier is $K = 2(C - 1)$.

- E-Measure: A measure of the PR (precision-recall) for an established number of results. And this measure only considers the first 32 models. The E-measure is defined as follows Eq. (4):

$$E = \frac{2}{\frac{1}{P} + \frac{1}{R}}.$$ (4)

The maximum score is 1.0, a higher score explains a better result.

- Discounted Cumulative Gain(DCG): A statistic of the weights that correct results in the front of the list more than correct results later in the last of the list, because some users do not take care of elements in the end of the list. The result list R is transformed to a vector $\mathbf{g}$, if the element g_i get value 1, the element r_i get a correct result, otherwise the value is 0. Discounted cumulative gain is defined as follows Eq. (5):

$$\mathrm{DCG}_i = \begin{cases} g_1 & i = 1, \\ \mathrm{DCG}_{i-1} + \frac{g_i}{\log_2 i} & \text{otherwise.} \end{cases}$$ (5)

Then the result is divided by the maximum DCG to get the final value and n is the number of the dataset.

$$\mathrm{DCG} = \frac{\mathrm{DCG}_n}{1 + \sum_{j=2}^{C} \frac{1}{\log_2 j}}.$$ (6)

In the matching algorithm, we use Hausdorff distance and hash binary encoding as the matching algorithm respectively. Comparisons with other algorithms are shown in Tab. I. DI-SIFT is the proposed algorithm. And DI-SIFT-haus-100 means the Hausdorff distance is used with 100 images. DI-SIFT-hash-200 means hash binary encoding is used with 200 images.

Algorithm	NN	FT	ST	E	DCG
SD[20]	0.23	0.10	0.17	0.05	0.53
CSI-SD[9]	0.50	0.13	0.20	0.07	0.56
SSC[7]	0.48	0.08	0.12	0.05	0.51
ZFCE-BoF[9]	0.55	0.17	0.19	0.09	0.59
SC-GALIF[5]	0.63	0.28	0.38	0.15	0.68
DI-SIFT-haus-100	0.82	0.37	0.49	0.20	0.75
DI-SIFT-haus-200	0.83	0.38	0.50	0.20	0.75
DI-SIFT-hash-100	0.81	0.35	0.46	0.18	0.73
DI-SIFT-hash-200	0.60	0.41	0.54	0.21	0.74

Tab. I *The 5 evaluation criteria results of the SIFT algorithm.*

From Tab. I, we can see that the Hausdorff distance and hash binary matching get the similar result. When the number of projection image is 200, the matching result is slightly better than that of the number 100.
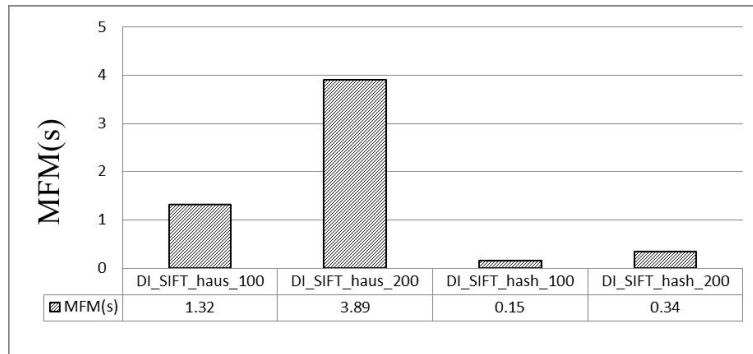
The P-R result of the DI-SIFT based on Hausdorff distance and hash binary encoding is given in Fig. 7. There is little difference between using Hausdorff distance and hash binary encoding algorithm. When projection images are 200, the hash binary encoding matching algorithm achieve the best result.
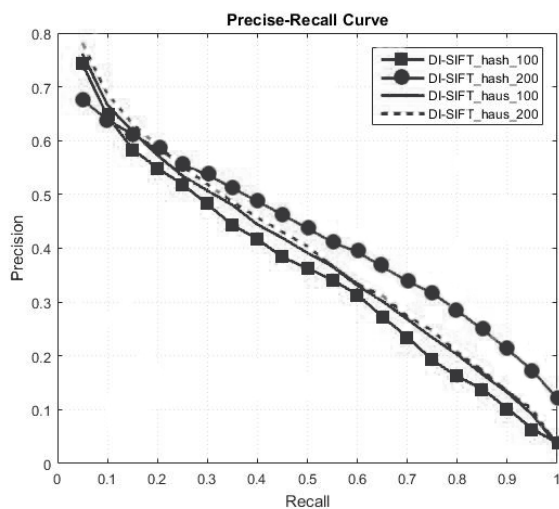

Fig. 7 *The P-R result of the DI-SIFT based on Hausdorff distance and hash binary encoding.*

Here we compare the time complexity of proposed method to DI-SIFT based on Hausdorff distance methods. A comparative experiment based on the Hausdorff distance and hash binary encoding of the average time (Mean Feature Match, (MFM) is shown in Fig. 8. From Fig. 8, we can find the matching time of the hash binary encoding is tens time faster than using Hausdorff distance.

Fig. 8 *The matching time comparison between using Hausdorff distance and hash binary encoding.*
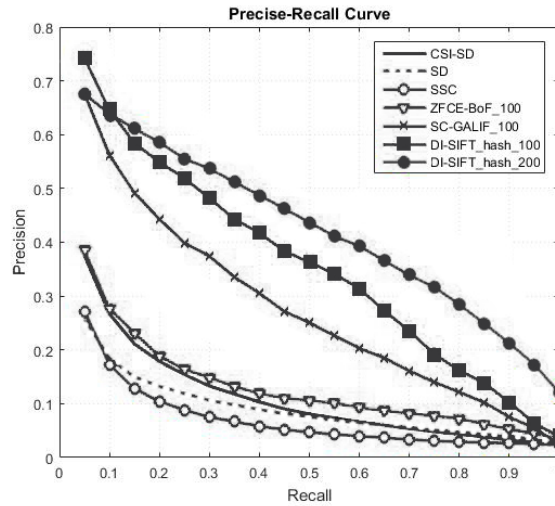
Fig. 9 *The P-R result of the DI-SIFT algorithm and other algorithm.*

The Precision-Recall curves of DI-SIFT algorithm and other algorithms are given in Fig. 9. Compared with SC-GALIF, ZFCE-BoF, SD, CSI-SD and SSC algorithm, DI-SIFT algorithm gets the best retrieval result.

5.   Conclusions

In this paper, we proposed a multi-view 3D model feature matching algorithm based on cascade hash binary encoding. In the processing of the feature extraction, we employed the laplacian operator to sharpen depth image and RANSAC as the filter. In the feature matching stage, the hash binary encoding achieved a similar result with the method based on Hausdorff distance with less matching time. However,

the computation cost of feature extraction is still high. Our future work will focus on improving the efficiency of multi-view 3D model feature extraction algorithm.

Acknowledgement

References

[1] ANSARY T.F., DAOUDI M., VANDEBORRE J.P. A bayesian 3-d search engine using adaptive views clustering. *IEEE Transactions on Multimedia*. 2007, 9(1), pp. 78–88, doi: 10.1109/TMM.2006.886359.

[2] AYADI M.G., BOUSLIMI R., AKAICHI J.A. medical image retrieval scheme through a medical social network. *Netw Model Anal Health Inform Bioinforma*. 2016, 5(23), doi: 10.1007/s13721-016-0130-9.

[3] CHAOUCH M., VERROUST-BLONDET A. A new descriptor for 2D depth image indexing and 3D model retrieval. In: *IEEE International Conference on Image Processing (ICIP 2007)*, pp. VI-373-VI-376.

[4] CHEN D., TIAN X., SHEN Y. On visual similarity based 3D model retrieval. In: *Computer graphics forum*, Blackwell Publishing, Inc., 2003, pp. 223–232.

[5] EITZ M., RICHTER R., BOUBEKEUR T., HILDEBRAND K., ALEXA M. Sketch-based shape retrieval. *ACM Transactions on Graphics (TOG)*. 2012, 31(4), doi: 10.1145/2185520.2185527.

[6] FISCHLER M.A., BOLLES R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*. 1981, 24(6), pp. 381–395, doi: 10.1145/358669.358692.

[7] FROME A., HUBER D., KOLLURI R. Recognizing objects in range data using regional point descriptors. In: *European conference on computer vision (ECCV 2004)*, Springer Berlin Heidelberg, 2004, pp. 224–237.

[8] GAO Y., DAI Q., ZHANG N. 3D model comparison using spatial structure circular descriptor. *Pattern Recognition*. 2010, 43(3), pp. 1142–1151, doi: 10.1016/j.patcog.2009.07.012.

[9] GAO Y., LIU A., NIE W., SU Y., DAI Q., CHEN F., CHEN Y., CHENG Y., DONG S., DUAN X., FU J., GAO Z., GUO H., GUO X., HUANG K., JI R., JIANG Y., LI H., LU H., SONG J., SUN J., TAN T., WANG J., YIN H., ZHANG C., ZHANG G., ZHANG Y., ZHAO C., ZHAO X., ZHU G. SHREC'15 Track: 3D Object Retrieval with Multimodal Views. In: *Eurographics Association*, 2015, pp. 129–136.

[10] KNOPP J., PRASAD M., WILLEMS G., TIMOFTE R., VAN GOOL L. Hough transform and 3D SURF for robust three dimensional classification. In: *Europan Conference on Computer Vision 2010 (ECCV 2010)*, Springer Berlin Heidelberg, 2010, pp. 589–602.

[11] LI B., LU Y., LI C., GODIL A., SCHRECK T., AONO M., CHEN Q., CHOWDHURY N.K., FANG B., FURUYA T., JOHAN H., KOSAKA R., KOYANAGI H., OHBUCHI R., TATSUMA A. SHREC' 14 Track: Large Scale Comprehensive 3D Shape Retrieval. In: *Eurographics Workshop on 3D Object Retrieval 2014 (3DOR 2014)*, 2014, pp. 131–140.

[12] LI D. An Artistic Rendering Framework for Stereoscopic Images. In: *Chinese Journal of computers*. 2014, 37(10), pp. 2218–2226. Available from: http://cjc.ict.ac.cn/online/onlinepaper/ldj-2014107105150.pdf

[13] LINDSTROM P., TURK G. Image-driven simplification. *ACM Transactions on Graphics (TOG)*. 2000, 19(3), pp. 204–241, doi: 10.1145/353981.353995.

[14] LIU L., WANG R., LUO X., FU X., LIU L. Data-Driven method for rapid 3D garment modeling. *Journal of Software*. 2016, 27(10), pp. 2574–2586. Available from: `http://www.jos.org.cn/1000-9825/5071.htm`

[15] LOWE D.G. Object recognition from local scale-invariant features In: *Proc. 7th International Conference on Computer Vision (ICCV)*, 1999, pp. 1150–1157.

[16] LOWE D.G. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision*. 2004, 60, pp. 91–110, doi: `10.1023/B:VISI.0000029664.99615.94`.

[17] MAHMOUDI S., DAOUDI M. 3D models retrieval by using characteristic views. In: *Proceedings of the 16th International Conference on Pattern Recognition*, IEEE, 2002, pp. 457–460.

[18] OHBUCHI R., OSADA K., FURUYA T., BANNO T. Salient local visual features for shape-based 3D model retrieval. In: *SMI 2008 IEEE International Conference on Shape Modeling and Applications*, IEEE, 2008, pp. 93–102.

[19] OHBUCHI R., FURUYA T. Distance metric learning and feature combination for shape-based 3D model retrieval. In: *Proceedings of the ACM workshop on 3D object retrieval*, ACM, 2010, pp. 63–68.

[20] OSADA R., FUNKHOUSER T., CHAZELLE B., DOBKIN D. Matching 3D models with shape distributions. In: *SMI 2001 International Conference on Shape Modeling and Applications*, IEEE, 2001, pp. 154–166.

[21] PAPADAKIS P., PRATIKAKIS I., THEOHARIS T., PERANTONIS S. PANORAMA: A 3D shape descriptor based on panoramic views for unsupervised 3D object retrieval. *International Journal of Computer Vision*. 2010, 89(2-3), 177-192. doi: `10.1007/s11263-009-0281-6`.

[22] TANGELDER J.W., VELTKAMP R.C. A survey of content based 3D shape retrieval methods. *Multimedia tools and applications*. 2008, 39(4), pp. 441–471 doi: `10.1007/s11042-007-0181-0`.

[23] TOLDO R., CASTELLANI U., FUSIELLO A. Visual vocabulary signature for 3D object retrieval and partial matching. In: *Proceedings of the 2nd Eurographics conference on 3D Object Retrieval*, Eurographics Association, 2009, pp. 21–28.

[24] VANAMALI T.P., GODIL A., DUTAGACI H., FURUYA T., LIAN Z., OHBUCHI R. SHREC'10 track: generic 3D warehouse. In: *Proceedings of the 3rd Eurographics conference on 3D Object Retrieval*, Eurographics Association, 2010, pp. 093–100.

[25] VRANIĆ D.V., SAUPE D. 3D model retrieval. *Proc. SCCG*. 2004, pp. 3–6.

[26] WANG M., GAO Y., LU K., RUI Y. View-based discriminative probabilistic modeling for 3d object retrieval and recognition. *Transactions on Image Processing*. IEEE, 2013, 22(4), pp. 1395–1407, doi: `10.1109/TIP.2012.2231088`.

[27] ZHOU P., YUAN J., LIU X., ZHU T. Three-Dimensional Extraction and Quantification of Facial Wrinkles *Journal of Computer-Aided Design and Computer Graphics*. 2015, 27(1), pp. 81–87. Available from: `https://www.researchgate.net/profile/Ping_Zhou15/publication/282375892_Three-dimensional_extraction_and_quantification_of_facial_wrinkles/links/56e0df2c08ae979addf0ffdc.pdf`