HAL
open science

A new descriptor for 2D depth image indexing and 3D model retrieval

Mohamed Chaouch, Anne Verroust-Blondet

To cite this version:

Mohamed Chaouch, Anne Verroust-Blondet. A new descriptor for 2D depth image indexing and 3D model retrieval. ICIP 2007 - International Conference on Image Processing, Sep 2007, San Antonio, United States. pp.VI - 373 - VI - 376, 10.1109/ICME.2007.4284721 . hal-00802581

HAL Id: hal-00802581

https://inria.hal.science/hal-00802581v1

Submitted on 20 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A NEW DESCRIPTOR FOR 2D DEPTH IMAGE INDEXING AND 3D MODEL RETRIEVAL

Mohamed Chaouch and Anne Verroust-Blondet

INRIA Rocquencourt Domaine de Voluceau, B.P. 105 78153 Le Chesnay Cedex, FRANCE
Email: {mohamed.chaouch, anne.verroust}@inria.fr

ABSTRACT

We present here a new descriptor for depth images adapted to 2D/3D model matching and retrieving. We propose a representation of a 3D model by 20 depth images rendered from the vertices of a regular dodecahedron. One depth image of a 3D model is associated to a set of depth lines which will be afterward transformed into sequences. The depth sequence information provides a more accurate description of 3D shape boundaries than using other 2D shape descriptors. Similarity computing is performed when dynamic programming distance (DPD) is used to compare the depth line descriptors. The DPD leads to an accurate matching of sequences even in the presence of local shifting on the shape. Results on a large 3D database show efficiency of our 2D/3D approach.

Index Terms— Depth image indexing, depth sequence, 2D/3D description, Content-based 3D model retrieval

1. INTRODUCTION

Content-based retrieval of 3D models has become an important subject of research (cf. a very good survey [1] and comprehensive comparative studies of 3D retrieval algorithms, reported in [2, 3, 4, 5]). The 2D/3D shape retrieval methods are based on the computation of 2D images rendered from multiple viewpoints, considering that two models are similar when they look similar from all viewing angles. The 3D model is then indirectly represented by various 2D-shape descriptors associated with 2D views so that the 3D-shape matching is transformed into similarity measuring between 2D images. Silhouettes and depth-buffer images are the most widespread projections. A silhouette is a binary image whereas a depth-buffer image contains the information of the distance between the object and the viewing plane in its pixels.

In [6], the silhouettes are encoded by their Zernike moments and Fourier descriptors. The dissimilarity between two 3D models is defined as the minimal dissimilarity of the silhouettes over all rotations and all pairs of vertices on the corresponding dodecahedrons. Filali Ansary et al [7] introduce a novel probabilistic Bayesian using a characteristic views selection. Curvature scale space (CSS) descriptor is used in [8].

This work has been supported in part by the DELOS NoE on Digital Libraries (EU IST NoE G038-507618

Most of the 2D/3D approaches based on depth images [9, 10, 11] use the two-dimensional discrete Fourier transform (2D-DFT) as a 2D-shape signature. They differ mainly in the similarity estimation technique used and in the number of views retained. Recently, Chaouch et al [10] have introduced a technique to enhance 2D/3D shape descriptors. To take into account the dispersion of information in the views, they associate to each view a relevance index. This approach can be applied to any type of 2D/3D descriptors, in particular to the Silhouette-based and the Depth buffer-based descriptors using the Fourier-transform, presented by Vranic [11]. In fact, the overall performance of a 2D/3D retrieval algorithm could also be increased by introducing a relevance feedback procedure [12], but our purpose here is to improve the efficiency of 2D/3D shape descriptors.

Our 2D/3D approach is based on depth images but differs with the previous ones. We introduce a new descriptor for depth images and a new similarity measure: first, we transform the lines of the depth image in order to represent it by a collection of sequences without loss of information; secondly, we use a Dynamic Programming distance contrarily to the previous techniques that adopt generally the Euclidean distance. In Section 2, we first present the depth line descriptor and the depth-image matching with dynamic programming. In Section 3, we describe the proposed 2D/3D shape retrieval method. Experimental results are given in Section 4 and we conclude in Section 5.

2. A NEW DESCRIPTOR FOR A DEPTH IMAGE

The proposed descriptor is essentially used for depth image indexing and is not adapted to natural images. To build it, our system generates $2 \times N$ depth lines, considering the depth image of size $N \times N$ as a collection of N horizontal and N vertical depth lines. Each depth line is represented by a sequence of states using our proposed sequencing method (see Fig. 1). A depth line descriptor consists in the set of $2 \times N$ sequences. The sequence extraction process is described in details in the following.

2.1. Sequencing of Depth Line

The challenge of the sequencing method lies in the definition of states, which need to be efficient and robust representations of the shape information. It is possible to assimilate

depth lines to 2D curves, thus the space points are taken as sequence elements. However, this sequencing technique has some disadvantages. When the depth line has several parts (e.g., contains disconnected background regions), there is a bad resolution in the matching process due to the fact that the dynamic programming algorithm matches empty samples. It is then interesting to discover, analyze and preserve the content of a depth line.

The different regions present on a depth line are classified into two categories:

Background regions: they represent the background of the projected plane. Sometimes, a 3D model might be rendered into several separated parts and also certain parts might contain holes. When the situation occurs, two types of background points can be rendered: interior points and exterior points. We want to distinguish these two background states in our sequencing.

Projected object regions: they contain the projections of the visible points of 3D model. For these regions, we adopt a "derivative" description that does not consider the depths of the visible points or "the attributes of pixels", but rather considers the first derivative of the sequences. While there are sophisticated methods for estimating derivatives, we use the following estimate for simplicity $D = dl(p) - dl(p+1)$. This is the slope of the line segment between the depth at p and the depth at its right neighbor. To represent this estimate at each point, we introduce three states, Decreased, Increased and Constant, according to the sign of its derivative (negative or positive or null). We obtain information about the shape by considering the space derivative properties of the depth line.

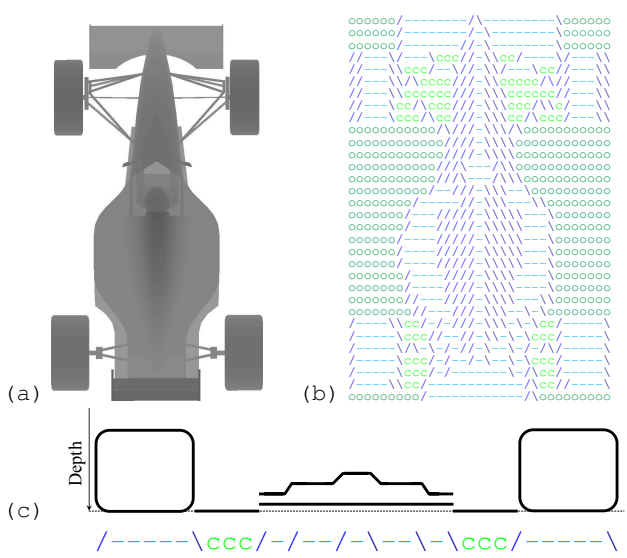


Fig. 1. (a). Depth image, (b). The horizontal sequences computed from the depth lines, (c). Sequencing of a depth line.

Each depth line is encoded in a set of N states called sequence of observations. Each such observation is represented by one of the five characters, o, c, /, -, \, representing the proposed states that compose the sequence: exterior-background, interior-background, increased-depth, constant-depth and decreased-depth (see Fig. 1.c).

2.2. Depth Image Matching Using Dynamic programming

For some sequence similarity matching, a very simple measure, such as Hamming distance, will suffice. It is the number of corresponding state positions that differ. However, this metric turned out not to be optimal in our sequence similarity computing. It is essentially due to the fact that two similar sequences often have approximately the same overall component shapes, but these shapes do not line up in X-axis. Thus, it is necessary to find the best alignment between two sequences before computing their similarity.

As one can see on Fig. 2, the Hamming distance does not tolerate phase shifting of subsequences of the shape, thus, similar shapes that are not perfectly aligned along the X-axis lead to counter intuitive high distance values. To overcome the limits of the Hamming metric, we introduce the Dynamic Programming distance to compare the shape descriptors, thus local deformations of objects' boundaries can be matched.

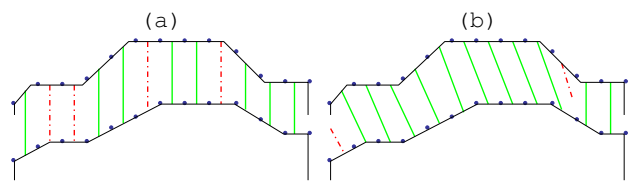


Fig. 2. Comparing the distances of two depth sequences by Hamming distance and dynamic programming distance while the sequences have an overall similar shape, but, they are not perfectly aligned along the X-axis. (a) 4 mismatches -·-·-, Hamming distance produces a pessimistic dissimilarity measure. (b) 2 mismatches -·-·-, a nonlinear alignment allows a more intuitive distance measure to be calculated.

2.2.1. *Dynamic Programming*

In diverse domains, the dynamic programming has been developed to align two signals, two sequences or two series with different dynamics. While there are many algorithms in the literature, Dynamic Time Warping (DTW), Needleman-Wunsch algorithm and Smith-Waterman algorithm account for the majority of the literature. DTW has been well-adopted for curve alignment [13] and 2D shape matching [14].

The Needleman-Wunsch algorithm [15], which calculates a "global" similarity distance between two sequences, is used here. It has already proven to yield superior performance for the retrieval of sequences, and we show that this is also the case for depth image matching (see [16] for more details).

2.2.2. *Measuring Similarity between two Depth Images*

To compare two depth images $\mathbf{I^1}$ and $\mathbf{I^2}$, we generate, for each one, $2 \times N$ observation sequences $\mathbf{S^1_{i,j}}$ and $\mathbf{S^2_{i,j}}$. Then, we compute the dynamic programming distance (DPD) between each sequence of image $\mathbf{I^1}$ and the corresponding se-

quence of image $\mathbf{I^2}$. Finally, the dissimilarity between $\mathbf{I^1}$ and $\mathbf{I^2}$, are computed as the sum of all DPDs. Therefore, it is defined as follows:

$$d(\mathbf{I^1}, \mathbf{I^2}) = \sum_{i=r,c} \sum_{j=0}^{N-1} DPD(\mathbf{S^1_{i,j}}, \mathbf{S^2_{i,j}}). \quad (1)$$

where i denotes if the depth line is a row or a column, and j denotes orders of lines in the depth image.

3. 3D MODEL RETRIEVAL

Our shape retrieval system compares 3D models based on their visual similarity using depth line descriptors extracted from depth images. The process first normalizes and scales each 3D model into a regular dodecahedron. Then, it generates the set of $N \times N$ depth images associated to the twenty vertices of the dodecahedron. Finally, it computes a depth line descriptor for each image. To sum up, twenty depth line descriptors will define the shape descriptor of the 3D model. The retrieval process is described step by step in the following.

3.1. Pose Normalization

A 3D object is generally given in arbitrary orientation, scale and position in the 3D space. As most of the 2D/3D shape descriptors do not satisfy the geometrical invariance, pose normalization is then necessary before the 3D object feature extraction. Some of the previous methods employed the Principal Component Analysis (PCA) as normalization step. We have retained here the "Continuous" PCA [17] (CPCA) because it appears to be more complete and the most stable of all the approaches we have studied.

3.2. Twenty Depth Line descriptors for representing a 3D model

Each 3D model is projected in the normal direction determined by the vertices of the regular dodecahedron, in order to generate twenty $N \times N$ depth images $\mathbf{I_i}$, $i \in \{1, ..., 20\}$. The rectangular images are deduced from the principal axes. Using our novel approach described in Section 2, our system extract 20 depth line descriptors from the generated images.

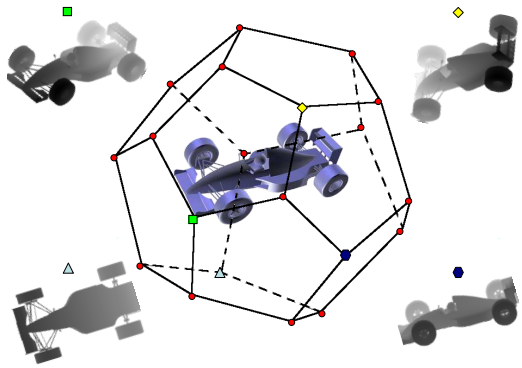


Fig. 3. Twenty viewpoints associated to dodecahedron's vertices, distributed uniformly over a race car model.

3.3. Measuring Similarity between two 3D Models

The similarity between two 3D models, $O1$ and $O2$, is measured here by summing up the similarity from all corresponding images as follows:

$$\Delta(O^1, O^2) = \sum_{i=1}^{20} d(\mathbf{I^1_i}, \mathbf{I^2_i}). \quad (2)$$

where $d(\mathbf{I^1_i}, \mathbf{I^2_i})$ is defined in Equation (1), and i denotes the index of the rendered images.

4. EXPERIMENTAL RESULTS

We made our tests on the Test Princeton 3D Shape Benchmark database [2] (907 models categorized within 92 distinct classes) with depth images of size 32x32. To compare objectively the retrieval effectiveness of the proposed approaches, we compute Precision-Recall diagrams commonly used in information search (the query is not counted in the answer as in [11]) and four quantitative measures for evaluating query results (see [2] for a description of this measures): **(1)** The Nearest Neighbor (NN), **(2)** The First Tier (FT), **(3)** the Second Tier (ST), **(4)** the Discounted Cumulative Gain (DCG).

In Fig. 5, we compare our novel Multi-view approach (MDLA) against the initial depth line-based approach (DLA) presented in [16]. Our enhanced method clearly outperforms the DLA: when using the Hamming distance (see Fig. 5.a), MDLA provides more accurate retrieval results than the DLA-HD and has a performance similar to the DLA-DPD, and by introducing the dynamic programming distance (see Fig. 5.b), the performance gain is increased. All these retrieval improvements are the result of the best covering of 3D shape information (20 views distributed uniformly over a 3D model). Moreover, if we compare these measures with the results listed in table 4 of [2], our approach produces 3%, 5.5%, 5.5% and 3.5% better than the top ranked Light Field Descriptor (LFD) [6], in term of NN, FT, ST and DCG measures.

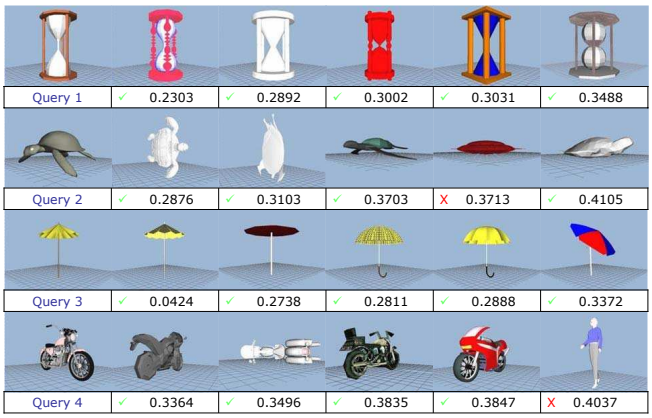
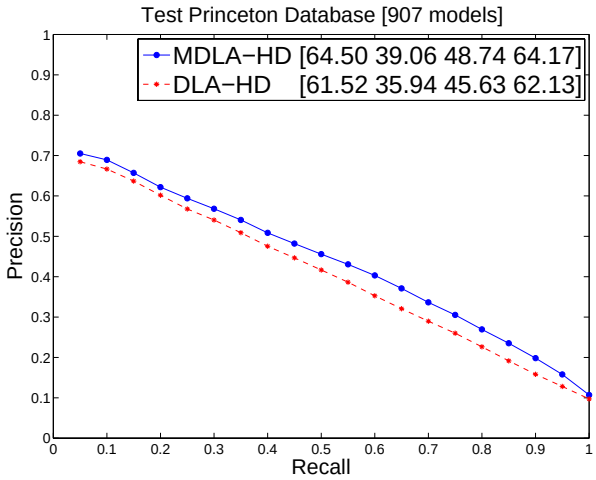


Fig. 4. Four queries from different classes of size 6. For each query, we show the top 5 objects matched with MDLA approach. The similarities between the query models and the retrieved models are given below corresponding images. ✓ and x indicate that the retrieved models belong or don't belong to the query's class, respectively.

5.a. Hamming distance HD



5.b. Dynamic programming distance DPD

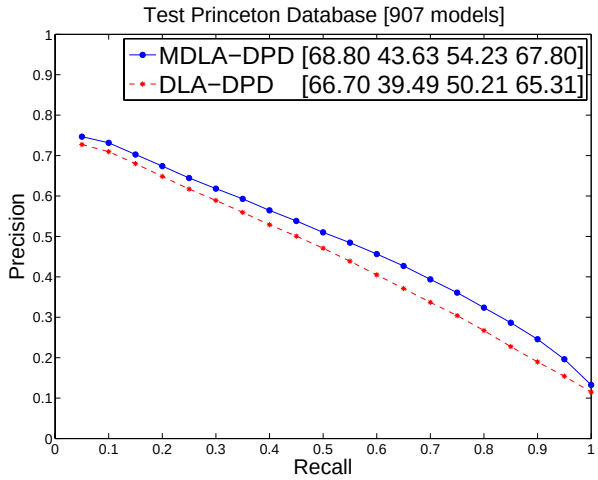


Fig. 5. Average Precision-recall curves using the proposed Multi-view depth line-based approach (MDLA, 20 views) and the initial approach (DLA[16], 6views). The mean NN, FT, ST and DCG values are given in the legends.

5. CONCLUSION

In this paper, the depth line descriptor and the associated similarity measure adopted in our 3D model retrieval method have been presented. By taking twenty depth line descriptors instead of six as in [16], we have increased its performances on the Princeton database. The depth-line approach has proved to be efficient in space and accuracy for 3D retrieval, even if it used with the Hamming distance. Work in progress includes the test of other dynamic programming algorithms and other values of N to obtain an optimal ratio time/accuracy.

6. REFERENCES

[1] J.W.H. Tangelder and R.C. Veltkamp, "A survey of content based 3D shape retrieval methods," in *SMI'04*, Genova, Italy, June 2004, pp. 145–156.

[2] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape benchmark," in *SMI'04*, Genova, Italy, June 2004, pp. 167–178.

[3] T. Zaharia and F. Prêteux, "3D versus 2D/3D shape descriptors: A comparative study," in *SPIE Conf. on Image Processing: Algorithms and Systems*, San Jose, CA, USA, Jan. 2004, vol. 5298.

[4] B. Bustos, D. A. Keim, T. Schreck, and D. Vranic, "An experimental comparison of feature-based 3D retrieval methods," in *3DPVT'04*, Thessaloniki, Greece, Sept. 2004.

[5] A. Del Bimbo and P. Pala, "Content-based retrieval of 3D models," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, no. 1, 2006.

[6] D.Y. Chen, X.P. Tian, Y.T. Shen, and M. Ouhyoung, "On visual similarity based 3D model retrieval," *Computer graphics forum*, vol. 22, no. 3, pp. 223–232, Sept. 2003.

[7] T. Filali Ansary, M. Daoudi, and J-P. Vandeborre, "3D model retrieval based on adaptive views clustering," in *ICAPR'05*, Bath, UK, Aug. 2005.

[8] S. Mahmoudi and M. Daoudi, "3D models retrieval by using characteristic views," in *ICPR'02*, Québec, Canada, Aug. 2002, pp. 11–15.

[9] R. Ohbuchi, M. Nakazawa, and T. Takei, "Retrieving 3D shapes based on their appearance," in *MIR'03*, Berkeley, CA, USA, Nov. 2003.

[10] M. Chaouch and A. Verroust-Blondet, "Enhanced 2D/3D approaches based on relevance index for 3D-shape retrieval," in *SMI'06*, Matsushima, Japan, June 2006.

[11] D.V. Vranic, *3D Model Retrieval*, Ph.D. thesis, U. of Leipzig, 2004.

[12] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 28, no. 7, pp. 1,088–1,099, July 2006.

[13] K. Wang and T. Gasser, "Alignement of curves by dynamic time warping," *The Annals of Statistics*, vol. 25, no. 3, pp. 1251–1276, 1997.

[14] E. J. Keogh, L. Wei, X. Xi, S.-H. Lee, and M. Vlachos, "Lb_keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures.," in *VLDB*, 2006, pp. 882–893.

[15] S. Needleman and C. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443–453, 1970.

[16] M. Chaouch and A. Verroust-Blondet, "3D model retrieval based on depth line descriptor," in *IEEE International Conference on Multimedia & Expo (ICME'07)*, Beijing, China, July 2007.

[17] D. Vranic, D. Saupe, and J. Richter, "Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics," in *2001 Workshop Multimedia Signal Processing*, Cannes, France, Oct. 2001.