International Conference on Computer Analysis of Images and Patterns -- CAIP 2017
The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-64689-3_21

On the Use of the Tree Structure of Depth Levels for Comparing 3D Object Views

Fabio Bracci, Ulrich Hillenbrand, and Zoltan-Csaba Marton
Institute of Robotics and Mechatronics, German Aerospace Center (DLR),
Weßling, Germany
{fabio.bracci, ulrich.hillenbrand, zoltan.marton}@dlr.de

Michael H. F. Wilkinson
Johann Bernoulli Institute for Mathematics and Computer Science,
University of Groningen (RuG), Groningen, The Netherlands
michael.h.f.wilkinson@rug.nl

Abstract. Today the simple availability of 3D sensory data, the evolution of 3D representations, and their application to object recognition and scene analysis tasks promise to improve autonomy and flexibility of robots in several domains. However, there has been little research into what can be gained through the explicit inclusion of the structural relations between parts of objects when quantifying similarity of their shape, and hence for shape-based object category recognition. We propose a Mathematical Morphology inspired hierarchical decomposition of 3D object views into peak components at evenly spaced depth levels, casting the 3D shape similarity problem to a tree of more elementary similarity problems. The matching of these trees of peak components is here compared to matching the individual components through optimal and greedy assignment in a simple feature space, trying to find the maximum-weight-maximal-match assignments. The matching thus achieved provides a metric of total shape similarity between object views. The three matching strategies are evaluated and compared through the category recognition accuracy on objects from a public set of 3D models. It turns out that all three methods yield similar accuracy on the simple features we used, while the greedy method is fastest.

Keywords: 3D shape similarity, tree matching, mathematical morphology, object recognition, scene analysis

1 Introduction

Nowadays 3D sensing technologies provide an ever growing number of low-cost and reliable sensors like the popular Kinect® and the Xtion® in both first and second versions. The depth data produced by these devices come at a high frame rate and provide a dense representation of 3D surface geometry. Speed of processing and representational geometric quality are important aspects in robotics, which often makes 3D sensing and processing the procedure of choice for enhancing robot autonomy and adaptivity. In particular, object recognition and scene analysis are increasingly based upon analysis of 3D data, and both imply matching and comparing objects.

Naive pixel-based or 3D point-based global matching is intractable for the combinatorial explosion of computations. Hence, several methods were developed to compare object shape through an abstract representation, e.g. histograms of local or regional descriptors on keypoints, global descriptors, or graphs of parts and their relations. Apart from the latter approach, the vast majority of representations has not explicitly regarded the relations of parts or components of objects. By contrast, we are here specifically interested in the effects of those relations when included in a *structural* object representation.

One particular kind of structure is the object's topology; comparing objects then means comparing topologies, which can be formalized as a graph matching problem. However, graph matching in general can be costly, even unfeasible for practical problems, therefore approaches like [14] take approximations and simplifications. Efficient matching algorithms are known for the graph subclass of trees, which are often produced as a problem approximation, e.g., by minimum spanning trees. Here we propose a representation where the favorable tree structure comes out naturally in the exact formulation.

The shape of 3D objects is here represented in a view-based fashion. Range images of object views are decomposed into *peak components*, which are the height profiles over 2D cross-sections taken at evenly spaced depth levels from the sensor. Considering depth levels is very suitable for depth data and this decomposition yields the inclusion relationship between each pair of such peaks: the inclusion hierarchy. A tree structure thus naturally arises from the proposed decomposition, the representation by a 3D *shape tree*.

Individual peak components, that is, the vertices of the 3D shape tree, need to be described in some feature space. For the sake of this study, we have here used a very simple feature space with limited descriptive capacity, thus relying on the aggregation effect of simple components. A 3D shape metric can be obtained by computing and quantifying a match between trees or simply between the sets of components.

We quantify the contribution of the tree structure among the components to the matching accuracy achieved in shape-based category recognition by a nearest-neighbor classifier: matching is done using the tree-structural constraints and also with structureless assignment of the components. Elaborating and comparing the results is the main contribution made by this work.

While intuitively one might expect that an explicit regard for structural relations among object components should generally improve matching accuracy, it here turns out that it does not improve similarity-based category recognition in this particular setup. It seems hence worthwhile to always compare a structural matching against a structureless baseline, and the latter may suffice. The analysis of the results is presented in section 6.

2 Related work

A taxonomy of shape matching methods is proposed in [22]; there shape matching methods are subdivided into *feature* based, *graph* based and *geometry* based. Our method uses the ability of a tree to describe the relations among the elements, which qualifies it as a graph-based method. At the same time the target to be described is a range image of an object taken from a given viewpoint, which also puts it among the geometry-based methods. The proposed method also relies on features of peak components, which makes it feature-based too.

An inspiration to our work has been the depth decomposition proposed in [8]. They slice 3D CAD models in three orthogonal directions aligned with the object's bounding box, and the cross-sections produced by scanning through the object along those directions are collected from the front to the back. The 3D descriptor is made by a histogram computed by binning simple 2D slice features.

A popular histogram-based shape descriptor is the Viewpoint Feature Histogram (VFH) of [19]. For any patch of points a histogram of three angular values between all the couples of point normals is computed and extended with the angular values between the point normals and the central viewpoint direction. A fast approximated kNN is then used to classify the objects. The methods based on histograms as global object descriptors do not take into account any structure among the constituent elements whose features are accumulated.

As discussed in [7], perceptual organization should be captured using models that take account of the part structure of objects and capture the properties of 3D shapes. As argued for example by Huber [6], part-based detection has the advantage of generalizing to unknown instances of object types. While in [6] and for the part-based VFH (called CVFH) feature [1] objects need to be singulated first, other part-based approaches like [13] can efficiently detect objects in clutter. Recently, Richtsfeld *et al.* [18] presented a multi-level approach to fit planar or curved surfaces to over-segment parts, and then define inter-segment relations to decide if they should be merged or not, but the method performs best for merging touching segments and for convex shapes.

A graph-based method is proposed in [21] where objects are matched through their skeletons. The skeletonization is made from a voxelized object and the skeletal voxels are connected with a minimum spanning tree algorithm. The actual object matching is restricted to the nearest graphs by indexing. The matching is modeled as a modified maximum-cardinality-minimum-weight matching, and is computed with a recursive depth-first coarse-to-fine search.

In the VRML community [27] also proposed a graph-based method. They decompose 3D objects in concave patches by watershed, and formulate the object matching as a graph isomorphism between attributed graphs, a computationally difficult problem without polynomial time solutions. The authors tackle the problem by merging patches such that their number stays small.

Trees for object recognition have been used in [4,5] for human airways recognition. The authors restrict the modeled trees to planted trees with bifurcations and trifurcations; edge weights represent the distance among vertices. The node similarity is computed from the ratio of the length of the airway to the vertex, the branch point type, and structural similarity. The matching between their structures minimizes a tree-edit distance, where a set of allowed matching combinations is previously determined to limit the $\mathcal{NP}$-complexity.

An example of a histogram-based region descriptor method is the shape context proposed in [2]. The authors first align the objects, and then sample a number of points along the object contour with uniform spacing, and in each contour point they take a distribution of the relative position of the other contour points. Correspondences between two objects are found by solving an optimal assignment problem; the total weight is used to estimate the similarity between pairs of objects. This method was extended to 3D in [11] to a histogram of relative 3D point positions of all the other shape points of a patch. The object matching goes through the optimal assignment formulation as well.

Deep learning is a class of methods of growing popularity, see [15]. The methods proposed so far are not able to process megapixel-size images and require large amount of training data. An example of such methods is [26] where a global object representation is learned, while in [3] a patches-representation is learned across deformed shapes.

The structural-representation-based method here investigated exploits the descriptiveness of geometric features of the object's parts at multiple scales, used along with a tree's ability to represent the relationship among those parts, while gearing on efficient tree matchings. This method was chosen as a promising structured matching procedure for depth data, as it is different from generic graph methods that are mainly applied rather on a small scale and on abstract decompositions/labellings of the objects due to the computationally costly matching.

This approach is similar to the work done in [10] where attributed graphs are built on a triangulation of Harris corners; they match image graphs by optimal assignment and locally preserve structure through the heterogeneous Euclidean overlap metric, a metric jointly considering vertex attributes, vertex degree, and attributes of the incident vertices. We also draw on the work of [24], where they measure the total distance between two trees with the maximum-cardinality-maximum-weight tree isomorphism through a recursive descent which maximizes the total accumulated similarity for different root vertex choices. They evaluate their work on matching skeletal graphs of 2D images.

In our application of these ideas we borrow the image decomposition from the work done in [25]. This method is based on 2D Shape-Size Pattern Spectra, which are 2D histograms of shape (moment of inertia divided by squared area) and size (area) of the peak components of an image. They classify images trough decision trees of the image's pattern spectra. Connected components and peak components are Mathematical Morphology's concepts which have been proven as a valuable decomposition; in [20] we find an efficient method to extract and store these components.

3 Method

The structural representation we investigate in this study aims at capturing shape details as well as the coarse outline. The representation is built from range images of objects taken from any viewpoint, either by a range sensor or through rendering of a 3D model. The core idea is to represent a 3D shape by a hierarchy of range image peaks, split into their connected components, and hence yielding a tree structure.

We now explain in turn the decomposition of a range image into its peak components and their descriptor; the construction of the tree structure; the matching algorithm for the trees; and how a 3D shape metric is derived from a match between two trees. The baseline of a structureless matching we compare to is a straightforward assignment between individual peak components from two object views.

Peak Components and Descriptors. As input we have a range image, which is a raster of depth values $D(x, y)$ for pixel coordinates (x, y). This raster describes a height profile of an object from a given viewpoint, with isolines shrinking towards the observer. Occluded parts and sides of the object are not visible to the observer and hence no information about those is available.

Slicing or thresholding at any depth level h produces one or more peak components P_i^h, being the connected image components of depths h and lower, such that their union

$$\cup_i P_i^h = \{(x,y)\,|D\,(x,y) \leq h\} \ . \tag{1}$$

In other words, as a sequence of decreasing depth levels are used for thresholding, the generated peak components are the connected regions in the remaining depth image. These are nested into each other like the isolines at the different values for h; the largest component lies at the maximal depth and the smallest lies closest to the sensor.

We use the same features as in [25] to describe the peak component shape, namely area, elongation and entropy. Additionally we consider relative area (the ratio of the peak component's area and the root peak component's area) and normalized entropy (the ratio of entropy and the logarithm of the number of pixels per peak component); see section 4 for further motivation.

The individual peak components are compared by means of Euclidean distances in a metric space. In a space of descriptive features, similar elements are expected to be close to each other while dissimilar elements are expected to be far apart. We map the point-wise distance to a similarity measure

$$s(u,v) = \max(1 - \|u - v\|/r, 0) \ , \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm, u and v are descriptor vectors of peak components, and r is a cutoff radius. In particular, the similarity is set to zero for a distance larger than r.

3D Shape Tree. So far we have an unorganized collection of peak components; now we represent their mutual inclusion by building the hierarchy. This is achieved through the computation of the Max-Tree proposed in [20], a well known data structure in the Mathematical Morphology. The Max-Tree is an inclusion tree, which is a tree of nested connected components, where each component includes its children. The Max-Tree represents morphological changes across the depth in the object through branches of arbitrary degree and depth, and through the vertex descriptors. The Max-Tree is a special level-set method.

The Max-Tree computation is linear in the number of pixels. This is achieved through a recursive flood-fill procedure starting at the highest depth; each time a pixel of lower depth is found, the flooding continues at the new depth level. This way the tree vertices correspond to varying discrete depth changes, and we have a non-uniform sampling of depth levels. We overcome this with an expansion step where all the vertices with depth difference to the parent $\triangle h > 1$ are substituted with a small series of $\triangle h$ non-branching vertices of unit depth increase.

Structural and Structureless Matching. Once object views are transformed into trees of peak component descriptors, we need a matching and comparison scheme defining quantitatively a 3D shape similarity. General graph matching is costly and when done by edit-distance minimization is $\mathcal{NP}$ complete; tree matching instead is solved efficiently in polynomial time.

We define a tree as $T = (V, E)$ with vertices V and edges E. A matching of the vertices of two trees without regarding the edges can be obtained straightforwardly through optimal assignment of the vertices. Specifically, we look for a matching $\varphi \subset V_1 \times V_2$ that maximizes the total similarity

$$W(\varphi) = \sum_{(u,v) \in \varphi} s\left(u, v\right). \tag{3}$$

where φ represents the one-to-one vertex mapping between the two trees. This optimal assignment is one baseline of a structureless matching we compare to. The problem was originally solved by Kuhn [12] and Munkres [16]. The original algorithm has $\mathcal{O}\left(n^4\right)$ complexity for n nodes to match; instead we use the algorithm of Jonker and Volgenant proposed in [9] which has $\mathcal{O}\left(n^3\right)$ complexity.

In the literature greedy algorithms are known which iteratively couple labeled vertices. We extend this approach with similarity (2), so that the correspondence with the largest weight is selected and added to the matching; the involved vertices are removed from the trees afterwards; this assignment proccess is iterated over the complete set. The procedure is shown in algorithm 1: it requires a distance matrix between all the vertex pairs, which has quadratic time complexity, and a sorting step on it, which gives $\mathcal{O}\left(n^2 \log\left(n\right)\right)$ total time complexity.

The above matching methods are generic and flexible, but they disregard the tree hierarchies and should be considered as *structureless* matching. For instance, if $u \in T_1$ is matched to $v \in T_2$, the child u' of u might be matched to the parent v" of v, as there is no constraint modeled in the assignment. We consider also a tree matching algorithm proposed in [24], which is proven to have $\mathcal{O}\left(bn^3\right)$ complexity, where b is the maximum degree of the tree vertices. For any given pair of vertices, u in T_1 and v in T_2, the algorithm performs a recursive descent and accumulates the similarity between u and v plus the maximal similarity among all the pairs of children of u and v in the matching: $(u', v') \in Ch\left(u\right) \times Ch(v)$ where $Ch\left(\right)$ gives

Algorithm 1 Sequential Greedy Match for MWMM

$(w, W) =$ `SequentialGreedyMWMM`(V_1, V_2) (V_1,V_2 vertex sets)

if $|V_1| = 0 \vee |V_2| = 0$ **then**
  return $(0, \emptyset)$
else
  $(u, v) = \text{argmin}\{s(u, v) \mid u \in V_1,\ \ v \in V_2\}$
  $V_1' = V_1 \setminus \{u\}\,;\ V_2' = V_2 \setminus \{v\}$
  $(w', W') =$ `SequentialGreedyMWMM`(V_1', V_2')
  return $(s\,(u, v) + w', \{(u, v)\} \cup W')$

the vertex children. The algorithm finds the maximal tree matching by iterating on $(u, v) \in \{root\,(T_1)\} \times T_2$ and $(u, v) \in T_1 \times \{root\,(T_2)\}$ where $root\,()$ is the tree root. It needs a pairwise similarity $s\,(u, v)$. This procedure resembles counting the similar vertices found in a parallel descent of two trees. Such matching enforces inclusion relationships by construction and is therefore a *structural* matching.

3D Shape Similarity Metric. Now a (structural or structureless) tree matching is computed in order to find the corresponding parts of two trees representing two object views. Let φ_{12} be the optimal mapping found through optimal assignment, greedy assignment or structural matching, hence maximizing (3). The total similarity $W\,(\varphi_{12})$ accumulated through φ_{12} is not meaningful as such: generally matching two small trees produces a small total similarity while matching two large trees produces a large one because of the different number of involved vertices. In order to avoid this size bias effect, we weight $W\,(\varphi_{12})$ relative to the original tree sizes. For this purpose we apply the four metrics proposed in [23], with $|V_i|$ the number of vertices in V_i:

$$d_1\,(T_1, T_2) = \max\,(|V_1|\,, |V_2|) - W\,(\varphi_{12}) \tag{4}$$

$$d_2\,(T_1, T_2) = |V_1| + |V_2| - 2W\,(\varphi_{12}) \tag{5}$$

$$d_3\,(T_1, T_2) = 1 - W\,(\varphi_{12})\,/\max\,(|V_1|\,, |V_2|) \tag{6}$$

$$d_4\,(T_1, T_2) = 1 - W\,(\varphi_{12})\,/\,(|V_1| + |V_2| - W\,(\varphi_{12})) \tag{7}$$

4 Evaluation

We evaluate the structural and structureless matching variants of our trees in an experiment on shape-based category recognition. We are interested in the accuracy of 1NN classification using the shape metrics (4) through (7) for finding the nearest neighbor.

Our experimental setting is similar to the one used in [2]. We take10 objects each from 10 classes from the SHREC-2010 database [17], take four specific views of each object (front, top, elevated back diagonal and elevated left diagonal view) and render a total of 400 range images of 900×1200 pixels. Each image is rendered with the focal length of the Kinect® and with an object-to-viewpoint distance such that the object fits in and fills the image. This set of arbitrary views is meant to contain different and therefore discriminable representations of the chosen objects. Example images, one for each class, are shown in figure 1.

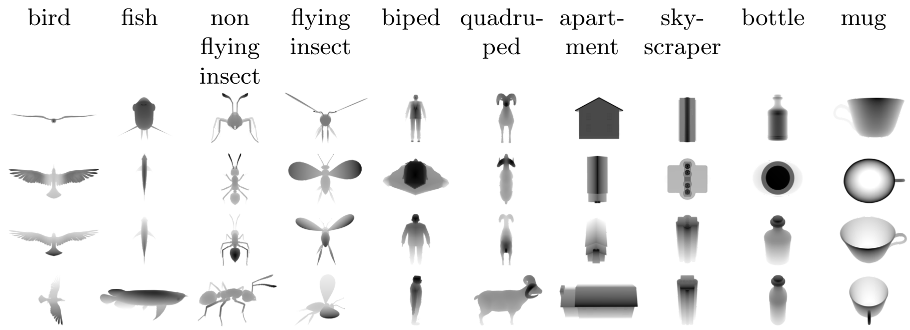

Fig. 1. Example range images from our evaluation set; the four views of an object from each class. Object models taken from [17].

In this study we focus on the aspect of shape similarity and disregard the absolute object size. Our 3D shape tree disregards absolute size by choice of the component descriptors. Moreover, to make the representation independent of the image size of an object as well, we investigate a version where the image area of a peak component is quantified as the fraction of the object silhouette area. Likewise, the entropy of a peak component is normalized by the logarithm of the component area. Finally, a depth normalization step is applied before the peak component decomposition in order to make the depth slicing comparable across objects. Depending on the actual application, when object size matters real physical units of the measures can be used for the descriptor vectors. Note that the used features are rotationally invariant which avoids the need for sampling of different angles around the viewing axis.

We compute the accuracy of a 1NN classifier of object category on three subsets of our range images set: a set of low-detail objects (fish, skyscraper, bottle, mug), a set of mixed-detail objects (bird, fish, biped, single house, mug), and a set of high-detail objects (bird, non flying insect, flying insect, biped, quadruped). This way we evaluate the quality of the proposed descriptor and relation between the different match procedures for different shape detail levels. Classification accuracies are computed in a leave-one-out scheme, where each queried object instance is left out from the database of category samples to match against. The evaluations cover the four shape metrics (4) through (7). The cutoff radius on similarity of peak components (cf. equation (2)) is set to a range of percentile values (10%, 20%,..., 90%, 99%) from the full sample of distances between all peak components. This way we adapt the parameter to the actual statistics of occurring feature distances.

5 Results

The evaluation was performed with four different features combinations; here we present results only for the feature set leading to the best classification accuracy, which is the relative area, normalized entropy and elongation. The results of the

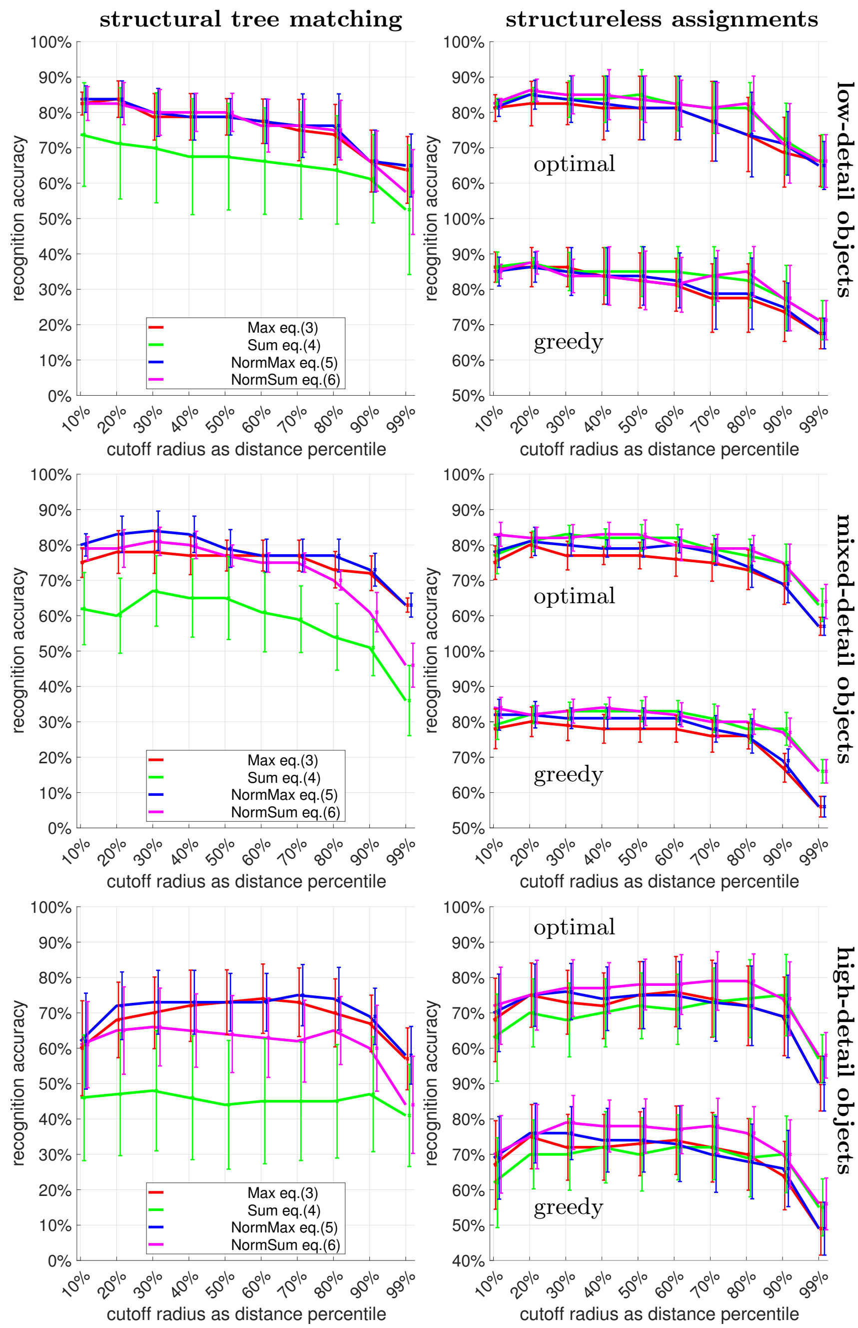

Fig. 2. Classification experiments results: mean classification accuracies and standard errors (bars) depending on similarity cutoff radius (cf. equation (2)). Shown is the performance separately on low-, mixed-, and high-detail shape classes for the structural and structureless matching methods and for the resulting shape metrics according to equations (4) through (7), as indicated.

classification experiments are shown in figure 2. There we can see the classification accuracy for matching the shape trees (structural match), and through optimal and greedy assignment (structureless matches) of peak components.

All the three matching methods show a similar best classification accuracy on the same data sets: on the low-detail shapes the greedy assignment is best with a 87.5% accuracy, on the mixed-details shapes both the greedy and the optimal assignments reach 84% accuracy, while on the high-detail shapes both the tree matching and the greedy assignment matchings show an 83% classification accuracy. The differences of best achievable accuracy between matching methods fall within the standard error of the mean in all these cases, are hence not significant. The overall best classification accuracy is achieved for the low-detail classes; it is overall worst for the high-detail classes.

The accuracy differences between best achievable accuracies fall within the standard error of the mean also across the different data sets. Nonetheless it's worth to note that the largest difference is between the accuracies on the low-detail and the high-detail shape classes, showing that low-detail shapes are somewhat easier to discriminate properly in the simple descriptor space considered. Very similar are the accuracies between different metrices within each data set: in most cases those are very close to each other with similar dependence on the different similarity cutoffs; only for the tree matching the Sum metric of equation (5) does have a lower accuracy. The general trend is that accuracies degrade for the highest similarity cutoffs, which indicates that there should be an indifferent zero similarity set from some moderate feature distance upward.

When comparing computation times for the different levels of shape detail, we found that low-detail classes are less computationally demanding than the high-detail classes for all the techniques. Also we observed an increasing computation time for both structureless matching variants when increasing the cutoff radius, while the structural matching remains insensitive to it.

6 Discussion and Conclusions

We have introduced a tree-structured representation of 3D object views that captures the hierarchy of peak components at different depth levels. The peak components have been described in a simple three-dimensional space of area, elongation, and entropy. In this setting, shape is represented by an aggregation of simple components. We have investigated the contribution made by the tree structure to the performance of a shape metric derived from matching these components between objects: matchings are computed with and without explicit regard for the edges in the tree as a structural constraint.

We have found no statistically significant difference in achievable accuracy of shape-based category recognition for the structural and structureless matches. When comparing the behavior for shape classes of different levels of geometric detail, it turnes out that low-detail shapes are somewhat easier to discriminate than high-detail shapes in our simple descriptor space. A very general conclusion to be drawn from this study is that it is worthwhile to compare any structural matching against a structureless baseline, such as the naive assignment. The latter may give similar results, be easier to implement, and may even run faster with off-the-shelf libraries.

It is interesting to understand the reason behind the somewhat counter-intuitive result that the tree structure of the shape components doesn't need to be considered during matching to find a reasonable correspondence. Our hypothesis is that this is due to the effect of the spatial arrangement of the peak components in the used feature space. On visual inspection it appears that their descriptors, as a point set in feature space, are generally laid out in the shape of the tree they are extracted from. Hence, two structurally and vertex-wise similar trees align quite well in feature space, such that the naive strategy of assigning components of the one tree to the nearest component of the other tree turns out effective and mostly does not disrupt the inclusion relationships.

As the found equivalence of structureless and structural assignments seems tightly linked to the nature of our feature space, it seems likely that results may turn out differently in different feature spaces. In particular, the role of structure in shape representations based on more descriptive features will be investigated in our research. Moreover, the structural aspect may be enhanced through, e.g., taking orientation relations between peak components into account.

References

1. Aldoma, A., Blodow, N., Gossow, D., Gedikli, S., Rusu, R., Vincze, M., Bradski, G.: CAD-Model Recognition and 6 DOF Pose Estimation Using 3D Cues. In: ICCV Workshop on 3D Representation and Recognition (3dRR11). Barcelona, Spain (2011)
2. Belongie, S., Malik, J., Puzicha, J.: Shape Context: A New Descriptor for Shape Matching and Object Recognition. In: Proceedings of the 13th International Conference on Neural Information Processing Systems. pp. 798–804. NIPS'00, MIT Press, Cambridge, MA, USA (2000),
3. Boscaini, D., Masci, J., Melzi, S., Bronstein, M.M., Castellani, U., Vandergheynst, P.: Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. Computer Graphics Forum 34(5), 13–23 (Aug 2015),
4. Graham, M.W., Higgins, W.E.: Optimal Graph-Theoretic Approach to 3D Anatomical Tree Matching. In: 3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano, 2006. pp. 109–112. IEEE (Apr 2006),
5. Graham, M.W., Higgins, W.E.: Globally optimal model-based matching of anatomical trees. In: Medical Imaging. vol. 6144, pp. 614415–614415–15 (2006),
6. Huber, D., Kapuria, A., Donamukkala, R.R., Hebert, M.: Parts-based 3d object classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 04) (July 2004)
7. Jacobs, D.W.: Perceptual organization as generic object recognition, vol. Volume 130, pp. 295 – 329. North-Holland (Dec 2001),
8. Jiantao, P., Yi, L., Guyu, X., Hongbin, Z., Weibin, L., Uehara, Y.: 3D Model Retrieval Based on 2D Slice Similarity Measurements. In: Proceedings of the 3D Data Processing, Visualization, and Transmission, 2Nd International Symposium. pp. 95–101. 3DPVT '04, IEEE Computer Society, Washington, DC, USA (2004),
9. Jonker, R., Volgenant, A.: A shortest augmenting path algorithm for dense and sparse linear assignment problems. Computing 38(4), 325–340 (Dec 1987),

10. Jouili, S., Mili, I., Tabbone, S.: Attributed Graph Matching Using Local Descriptions. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P. (eds.) Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science, vol. 5807, pp. 89–99. Springer Berlin Heidelberg (2009),
11. Körtgen, M., Novotni, M., Klein, R.: 3D shape matching with 3D shape contexts. In: In The 7th Central European Seminar on Computer Graphics (2003),
12. Kuhn, H.W.: The Hungarian method for the assignment problem. Naval Research Logistics 2(1-2), 83–97 (Mar 1955),
13. Lai, K., Fox, D.: Object recognition in 3d point clouds using web data and domain adaptation. The International Journal of Robotics Research 29(8), 1019–1037 (2010),
14. Marton, Z.C., Balint-Benczedi, F., Mozos, O., Blodow, N., Kanezaki, A., Goron, L., Pangercic, D., Beetz, M.: Part-Based Geometric Categorization and Object Reconstruction in Cluttered Table-Top Scenes. Journal of Intelligent & Robotic Systems 76(1), 35–56 (2014),
15. Masci, J., Rodolà, E., Boscaini, D., Bronstein, M.M., Li, H.: Geometric Deep Learning. In: SIGGRAPH ASIA 2016 Courses. SA '16, ACM, New York, NY, USA (2016),
16. Munkres, J.: Algorithms for the Assignment and Transportation Problems. Journal of the Society for Industrial and Applied Mathematics 5(1), 32–38 (1957),
17. Pratikakis, I., Spagnuolo, M., Theoharis, T., Editors, R.V., Dutagaci, H., Godil, A., Cheung, C.P., Furuya, T., Hillenbrand, U., Ohbuchi, R.: SHREC 2010 - Shape Retrieval Contest of Range Scans. In: Proceedings of Eurographics (2010),
18. Richtsfeld, A., Morwald, T., Prankl, J., Zillich, M., Vincze, M.: Segmentation of unknown objects in indoor environments. In: Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on. pp. 4791–4796 (2012)
19. Rusu, R.B., Bradski, G., Thibaux, R., Hsu, J.: Fast 3D recognition and pose using the Viewpoint Feature Histogram. In: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on. pp. 2155–2162. IEEE (Oct 2010),
20. Salembier, P., Serra, J.: Flat Zones Filtering, Connected Operators, and Filters by Reconstruction. IEEE Transactions on Image Processing 4(8), 1153–1160 (Aug 1995),
21. Sundar, H., Silver, D., Gagvani, N., Dickinson, S.: Skeleton based shape matching and retrieval. In: Shape Modeling International, 2003. pp. 130–139. IEEE (2003)
22. Tangelder, J.W.H., Veltkamp, R.C.: A survey of content based 3d shape retrieval methods. In: In Shape Modeling International. pp. 145–156 (2004),
23. Torsello, A., Hidovic, D., Pelillo, M.: Four metrics for efficiently comparing attributed trees. In: Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. vol. 2, pp. 467–470 Vol.2. IEEE (Aug 2004),
24. Torsello, A., Hidovic-Rowe, D., Pelillo, M.: Polynomial-time metrics for attributed trees. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27(7), 1087–1099 (Jul 2005),
25. Urbach, E.R., Roerdink, J.B.T.M., Wilkinson, M.H.F.: Connected Shape-Size Pattern Spectra for Rotation and Scale-Invariant Classification of Gray-Scale Images. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29(2), 272–285 (Feb 2007),
26. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D ShapeNets: A deep representation for volumetric shapes - IEEE Xplore Document. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Jun 2015),
27. Zuckerberger, E., Tal, A., Shlafman, S.: Polyhedral surface decomposition with applications. Computers & Graphics 26(5), 733–743 (Oct 2002),