Supplementary Material
for
A Graph Theoretic Approach for Object Shape Representation in Compositional Hierarchies using a Hybrid Generative-Descriptive Model

Umit Rusen Aktas[*], Mete Ozay[*], Aleš Leonardis and Jeremy L. Wyatt

School of Computer Science, The University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom.
Emails: { u.aktas, m.ozay, a.Leonardis, j.l.wyatt } @cs.bham.ac.uk

1   Introduction

In this supplementary material, we provide images of parts, compositions and realizations that are constructed and detected using Compositional Hierarchy of Parts (CHOP) algorithm which is introduced in the ECCV 2014 paper titled "A Graph Theoretic Approach for Object Shape Representation in Compositional Hierarchies using a Hybrid Generative-Descriptive Model".

Additionally, results which are obtained employing the CHOP on the images belonging to other datasets, and other images belonging to the datasets used in the ECCV 2014 paper, are given in addition to the ones that are provided in the paper.

2   Experiments

We examine our proposed approach on three benchmark object shape datasets, which are namely the Amsterdam Library of Object Images (ALOI) [1], the Tools and the Myth [2]. In the experiments [1], we used $\Theta = 6$ number of different orientations of Gabor features with the same Gabor kernel parameters implemented in [4]. We used subsampling ratio as $\sigma = 0.5$. In Section 2.1, we provide the results regarding shareability of parts that are constructed in the experiments presented in Section 4 of the main text of the ECCV 2014 paper. Section 2.2 includes experimental results on vocabulary learning with images of objects that are captured from multiple-views.

---

[*] The first and second author contributed equally.
[1] a Matlab implementation of CHOP is available on the webpage https://github.com/rusen/CHOP.git.

2.1   Results on Part Shareability

In this section, we analyze the effect of increasing number of objects, views, and categories on the degree of shareability of parts in the vocabulary. The shareability of a part in a learned vocabulary is measured as the average number of objects, views and categories that the part is shared across multiple objects belonging to a category, multiple views of an object belonging to a category, and multiple objects belonging to multiple categories, respectively. The shareability of a vocabulary is defined as the average shareability of its parts. In order to consider the shareability of the *most* descriptive parts in the analyses, we select 10 parts which have the best (lowest) MDL scores at each layer of the hierarchical vocabulary. Parts computed at the first layer $l = 1$ are not considered in shareability analyses, because Gabor filter responses calculated at the first layer are shared across almost all images in all of the experiments.

The relationship between the number of objects belonging to a single class and the shareability property of the shape model is illustrated in Fig. 1.a. The results show that the shareability of parts increases as the number of objects increases. The images used in this experiment are selected from the "Apple Logos" class in the ETHZ Shape Classes dataset [3]. The instances exhibit high variability in scale and rotation, therefore reducing part shareability. Similarly, in Fig. 1.b, we analyze the shareability of parts as the number of views of the same object increases. In this set of experiments, we selected a cup from Washington image dataset [1], which is used in Section 4.3 of the ECCV paper. Since similar parts are constructed using the images of objects captured at different views, we obtain higher shareability scores when compared to the single-category case in Fig. 1.a. Finally, Fig. 1.c illustrates the effect of training a vocabulary with objects belonging to different number of categories on the shareability of parts.
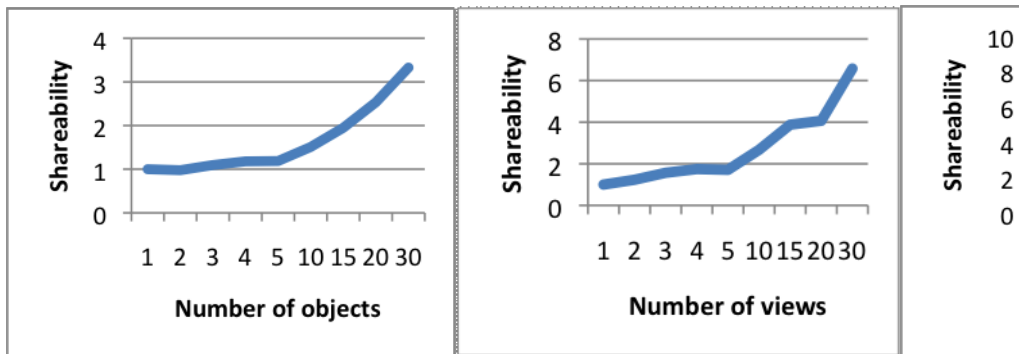
(a) Shareability vs Number of Objects   (b) Shareability vs Number of Views   (c) Shareal

Fig. 1: Shareability analysis of the three sets of experiments in ECCV 2014 paper.

2.2 Experiments on Multiple-View Images

Amsterdam Library of Object Images (ALOI) [6] dataset consists of multiple view images of objects belonging to 1000 categories. In the experiments presented in this section, we used 14 images captured from the viewpoints labelled $\{25^o, 50^o, 75^o, \ldots, 350^o\}$ as test images, and 14 images captured from the viewpoints labelled $\{30^o, 55^o, 80^o, \ldots, 355^o\}$ as training images, for each object.

In the first set of experiments, we analyzed the part shareability and computational complexity of the algorithms across multiple view images of a cup and a duck. For each layer $l = 1, 2, 3, 4, 5$, part realizations and object graphs detected on multiple view cup images and duck images are shown in Table 1 and Table 2, respectively. In the images, each part with a different part realization id is depicted by a different color. For instance, for an image of a cup captured from a viewpoint labelled as 75°, there are 6 different types of parts with 78 different part realizations at the first layer $l = 1$ (see second column of Table 1). However, we observe 5 different types of part compositions at the fifth layer $l = 5$ of the hierarchy. In the results, each node of an object graph, which is visualized by red points and lines, represents the position of the center of a part.

In the analyses of graph structures, we observe that the locality of topological structures of object graphs decreases through the higher layers representing object shapes with higher abstraction. For instance, part realizations of the parts represented with Gabor features at the first layer are connected to each other in a spatial neighbourhood in the results shown at $l = 1$ and $l = 2$ in Table 1 and Table 2. However, the connectivity of part realizations is determined using statistical and descriptive relationships between parts at the higher layers; horizontally oriented part realizations detected at the top and bottom of a cup and a duck are connected to each other, and vertically oriented part realizations detected at the right and left of the cup and duck are connected to each other for $l \geq 3$ in Table 1 and Table 2.

Table 1: Results on multiple view cup images obtained from ALOI Dataset

Rotation Degree	25°	75°	150°	225°	300°	350°
Original Image						
Layer $l = 1$						
Object Graph at $l = 1$						
Layer $l = 2$						
Object Graph at $l = 2$						
Layer $l = 3$						
Object Graph at $l = 3$						
Layer $l = 4$						
Object Graph at $l = 4$						
Layer $l = 5$						
Object Graph at $l = 5$						

Table 2: Results on multiple view duck images obtained from ALOI Dataset

Rotation Degree	25°	75°	150°	225°	300°	350°
Original Image						
Layer $l = 1$						
Object Graph at $l = 1$						
Layer $l = 2$						
Object Graph at $l = 2$						
Layer $l = 3$						
Object Graph at $l = 3$						
Layer $l = 4$						
Object Graph at $l = 4$						
Layer $l = 5$						
Object Graph at $l = 5$						
Layer $l = 6$						
Object Graph at $l = 6$						

2.3 Experiments on Partial Shape Similarity

Employing part shape similarity for learning compositions of parts is an important requirement for hierarchical compositional architectures [5]. In this section, we examine this property of the proposed CHOP algorithm in an articulated shape dataset called the Myth dataset [2].

In the Myth dataset, there are three categories, namely *Centaur*, *Horse* and *Man*. There are 5 different images belonging to 5 different objects in each category. Shapes observed in images differ by additional parts, e.g. the shapes of the objects belonging to Centaur and Man categories share the upper part of a man's body, and the shapes of the objects belonging to Centaur and Horse categories share the lower part of a horse's body. In the experiments, four samples belonging to each category is used for training and the other three images are used for testing. The results of four experiments are shown in Table 3, 4, 5 for Centaur, Horse and Man categories, respectively. The results are shown for the last two layers that are achieved in the construction of object graphs for each shape. In the tables, the right column labeled $l + 1$ represents the top layer, and the left column labeled l represents the previous column. For instance, the left column of Centaur-1 shape depicts part realizations and object graphs detected at the layer $l = 7$, and the right column depicts part realizations and object graphs detected at the layer $l + 1 = 8$ of the hierarchy in Table 3. Note that top layers of inference trees at which part realizations and object graphs are detected can be different for different shapes and images, since a hierarchical vocabulary and inference trees are dynamically constructed in the CHOP.

In the experiments, we first observe that the depths of inference trees of the objects belonging to the same category are closer to each other than those of the objects belonging to different categories. For instance, the depth of inference trees for three Centaur shapes are 8 and that of one Centaur shape is 7. Meanwhile, the depth of inference trees of three Man shapes are 6 and that of one Man shape is 7.

Moreover, we observe that the shared parts are correctly detected in part realizations and successfully employed in the construction of compositions. For instance, legs of horses which are shared among Centaur and Horse categories are represented as single compositions in the vocabularies and detected as realizations with unique ids at the top layer of the inference trees. However, back parts of horses are depicted with different shapes, therefore these parts are not shared across categories. Consequently, the unshared parts are not detected in the inference trees and are not of great significance in the construction of part vocabularies. Similarly, the articulated right arms of man shapes which are shared across five shapes belonging to Man and Centaur categories are detected in the inference trees.

Table 3: Results on images belonging to Centaur category obtained from Myth Dataset

Object Name, *Layer ID*	l		$l+1$	
	Part Realizations R^l	Object Graph $\mathbb{G}_l$	Part Realizations R^{l+1}	Object Graph $\mathbb{G}_{l+1}$
Centaur-1, $l = 7$				
Centaur-2, $l = 7$				
Centaur-3, $l = 7$				
Centaur-4, $l = 6$				

2.4 Experiments on Articulated Shape Images

In the last set of experiments, we examined the proposed approach using the articulated Tools dataset [2]. The dataset consists of 35 shapes belonging to 4 categories. Images belonging to Scissors and Pliers categories are used in the experiments. In each experiment, we selected one object belonging to a category as a training object and the other object in the same category as a test object. An articulation is used to construct different shapes of objects. Experiments on Scissors and Pliers categories are shown in Table 6 and 7, and Table 8 and 9, respectively. For instance, the images belonging to Scissors-2 are used for training a vocabulary of a CHOP for detection of parts of shapes in images belonging to Scissors-1 in the experiments given in Table 6, and vice versa in Table 7.

Table 4: Results on images belonging to Horse category obtained from Myth Dataset
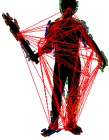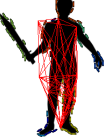
Object Name, *Layer ID*	l		$l+1$	
	Part Realizations R^l	Object Graph $\mathbb{G}_l$	Part Realizations R^{l+1}	Object Graph $\mathbb{G}_{l+1}$
Horse-1, $l=7$				
Horse-2, $l=7$				
Horse-3, $l=6$				
Horse-4, $l=6$				

Table 5: Results on images belonging to Man category obtained from Myth Dataset

Object Name, *Layer ID*	l		$l+1$	
	Part Realizations R^l	Object Graph $\mathbb{G}_l$	Part Realizations R^{l+1}	Object Graph $\mathbb{G}_{l+1}$
Man-1, $l = 5$				
Man-2, $l = 5$				
Man-3, $l = 5$				
Man-4, $l = 6$				

In the results, junctions and closed curves observed at the shape boundaries are detected as part realizations, if they are shared among different articulated images. Moreover, these shape parts are represented as single part compositions at the top layers of inference trees by object graphs. For instance, circular shape handles of scissors and V shaped handles of pliers are represented as compositions with unique ids in Table 6 and 7, and Table 8 and 9, respectively.

Table 6: Results on images of Scissor-1 object belonging to Scissor category obtained from Tools Dataset

Object Name, Articulation ID, *Layer ID*	l		$l+1$	
	Part Realizations R^l	Object Graph $\mathbb{G}_l$	Part Realizations R^{l+1}	Object Graph $\mathbb{G}_{l+1}$
Scissor-1, Art-1, $l=6$				
Scissor-1, Art-2, $l=6$				
Scissor-1, Art-3, $l=6$				
Scissor-1, Art-4, $l=5$				
Scissor-1, Art-5, $l=6$				

Table 7: Results on images of Scissor-2 object belonging to Scissor category obtained from Tools Dataset
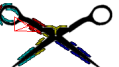
Object Name, Articulation ID, Layer ID	l		$l+1$	
	Part Realizations R^l	Object Graph $\mathbb{G}_l$	Part Realizations R^{l+1}	Object Graph $\mathbb{G}_{l+1}$
Scissor-2, Art-1, $l=5$				
Scissor-2, Art-2, $l=5$				
Scissor-2, Art-3, $l=5$				
Scissor-2, Art-4, $l=5$				
Scissor-2, Art-5, $l=5$				

Table 8: Results on images of Pliers-1 object belonging to Pliers category obtained from Tools Dataset
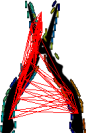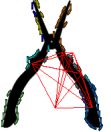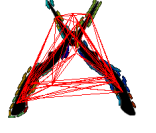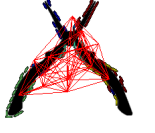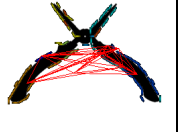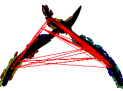
Object Name, Articulation ID *Layer ID*	l		$l+1$	
	Part Realizations R^l	Object Graph $\mathbb{G}_l$	Part Realizations R^{l+1}	Object Graph $\mathbb{G}_{l+1}$
Pliers-1, Art-1, $l=5$				
Pliers-1, Art-2, $l=4$				
Pliers-1, Art-3, $l=5$				
Pliers-1, Art-4, $l=5$				
Pliers-1, Art-5, $l=5$				

Table 9: Results on images of Pliers-2 object belonging to Pliers category obtained from Tools Dataset

Object Name, Articulation ID Layer ID	l		$l+1$	
	Part Realizations R^l	Object Graph $\mathbb{G}_l$	Part Realizations R^{l+1}	Object Graph $\mathbb{G}_{l+1}$
Pliers-2, Art-1, $l=5$				
Pliers-2, Art-2, $l=5$				
Pliers-2, Art-3, $l=5$				
Pliers-2, Art-4, $l=5$				
Pliers-2, Art-5, $l=5$				

References

1. L. Bo, K. Lai, X. Ren, and D. Fox, "Object recognition with hierarchical kernel descriptors," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '11.   Washington, DC, USA: IEEE Computer Society, 2011, pp. 1729–1736.
2. A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, and R. Kimmel, "Analysis of two-dimensional non-rigid shapes," *Int. J. Comput. Vision*, vol. 78, no. 1, pp. 67–88, Jun 2008.
3. V. Ferrari, T. Tuytelaars, and L. V. Gool, "Object detection by contour segment networks," in *Proceeding of the European Conference on Computer Vision*, ser. LNCS, vol. 3953.   Elsevier, June 2006, pp. 14–28.
4. S. Fidler and A. Leonardis, "Towards scalable representations of object categories: Learning a hierarchy of parts," in *Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ser. CVPR '07, June 2007, pp. 1–8.
5. S. Fidler, M. Boben, and A. Leonardis, "Learning hierarchical compositional representations of object structure," in *Object categorization computer and human perspectives*, S. J. Dickinson, A. Leonardis, B. Schiele, and M. J. Tarr, Eds. Cambridge, UK: Cambridge University Press, 2009, pp. 196–215.
6. J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005. [Online]. Available: http://www.science.uva.nl/research/publications/2005/GeusebroekIJCV2005