EG UK Computer Graphics & Visual Computing (2016)
Cagatay Turkay and Tao R. Wan (Editors)

Tactile Mesh Saliency: A Brief Synopsis

Manfred Lau and Kapil Dev
Lancaster University

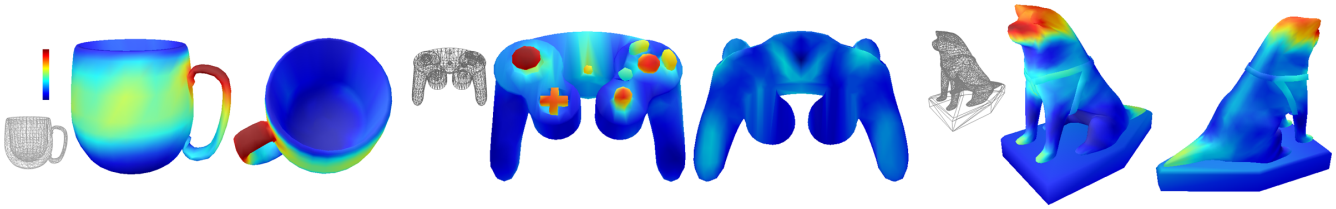

Figure 1: *Three examples of input 3D mesh and tactile saliency map (two views each) computed by our approach. Left: "Grasp" saliency map of a mug model. Middle: "Press" saliency map of a game controller model. Right: "Touch" saliency map of a statue model. The blue to red colors (jet colormap) correspond to relative saliency values where red is most salient.*

Abstract

*This work has previously been published [LDS*16] and this extended abstract provides a synopsis for further discussion at the UK CGVC 2016 conference. We introduce the concept of tactile mesh saliency, where tactile salient points on a virtual mesh are those that a human is more likely to grasp, press, or touch if the mesh were a real-world object. We solve the problem of taking as input a 3D mesh and computing the tactile saliency of every mesh vertex. The key to solving this problem is in a new formulation that combines deep learning and learning-to-rank methods to compute a tactile saliency measure. Finally, we discuss possibilities for future work.*

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Modeling Packages

1. Tactile Mesh Saliency: A New Concept

We have introduced the concept of tactile mesh saliency [LDS*16], and we describe this concept and further explorations for future work in this extended abstract. An important aspect of a geometric shape is its saliency, which are features that are more significant especially when comparing regions of the shape relative to their neighbors. The concept of *visual saliency* has been well studied in image processing [IKN98, BJB*15]. "Mesh Saliency" [LVJ05] is a closely related work that explores visual saliency for 3D meshes. However, other sensory stimuli have not been explored for mesh saliency. We introduce the concept of *tactile mesh saliency* and bring the problem of mesh saliency from the modality of visual appearances to tactile interactions. We imagine a virtual 3D model as a real-world object and consider its tactile characteristics.

We consider points on a virtual mesh to be tactile salient if they are likely to be grasped, pressed, or touched by a human hand. For our concept of tactile saliency, the human does not directly interact with real objects, but considers virtual meshes as if they were real objects and perceives how he/she will interact with them. We focus on a subset of three tactile interactions: grasp (*specifically for grasping to pick up an object*), press, and touch (*specifically for touching of statues*). For example, we may grasp the handle of a cup to pick it up, press the buttons on a mobile device, and touch a statue as a respectful gesture. The ideas of grasp synthesis for robots [SEkB12] and generation of robotic grasping locations [VPV12] have been explored in previous work. However, the existing work in these areas solve different problems and have different applications. The problem we solve in this paper is to take an input 3D mesh and compute the relative tactile saliency of all vertices on the mesh.

Computing tactile mesh saliency from geometry alone is a challenging, if not impossible, computational problem. Yet humans have great intuition at recognizing such saliency information for many 3D shapes. While a human finds it difficult to assign absolute saliency values (e.g. vertex i has value 0.8), he/she can typically rank whether one point is more tactile salient than another (e.g. vertex i is more likely to be grasped than vertex j). Hence we do not, for example, solve the problem with a regression approach. The human-provided rankings lead us to a ranking-based learning approach. However, recent similar learning approaches in graphics [GAGH14, LHLF15] typically learn simple scaled Euclidean distance functions. In contrast, we combine a deep architecture (which can represent complex non-linear functions) and a learning-to-rank method (which is needed for our "ranking"-based data) to develop a deep ranking formulation for the tactile mesh saliency problem and contribute a new backpropagation as the solution.

We first collect crowdsourced data where humans compare the tactile saliency of pairs of vertices on various 3D meshes (Figure 2). We represent a 3D shape with multiple depth images taken from different viewpoints. We take patches from the depth images and learn a deep neural network that maps a patch to a saliency value for the patch center (Figure 3). The same deep neural network can

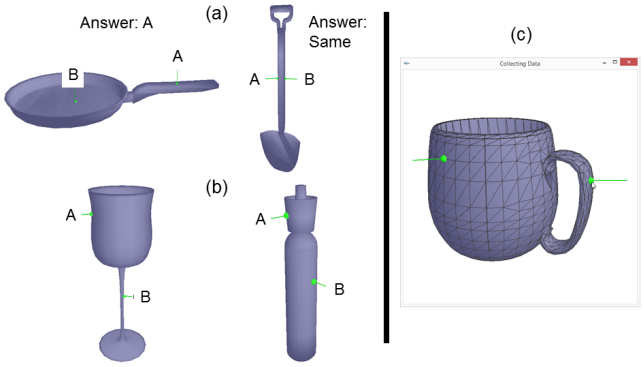

Figure 2: *(a) Two examples of images with correct answers given as part of the instructions for Amazon Mechanical Turk HITs. Text instructions were given to users: they are specifically asked to imagine the virtual shape as if it were a real-world object, and to choose which point is more salient (i.e. grasp to pick up, press, or touch for statue) compared to the other or that they have the same saliency. (b) Two examples of images of HITs we used. (c) Screenshot of software where user directly selects pairs of vertices and specify which is more salient (or same).*

be used across different depth images and 3D shapes, while different networks are needed for each tactile modality. After the learning process, we can take a new 3D mesh and compute a tactile saliency value for every mesh vertex. Since our approach is based on ranking, these are relative values and have more meaning when compared with each other. We compute saliency maps for three tactile interactions for 3D meshes from online sources including Trimble 3D Warehouse and the Princeton Shape Benchmark [SMKF04].

2. Further Explorations

Starting from our new concept of tactile mesh saliency, there can be many potential avenues for future work. First, we can experiment with more data types. This includes other tactile modalities and other possible types of human interactions with virtual and real objects. We have collected data on user perceptions of interactions with virtual 3D meshes. In the future, we can also collect data where humans interact with real-world objects.

Second, there is more to deep learning that we can explore. We have leveraged two fundamental strengths of deep learning by having an architecture with multiple layers and by not using handcrafted 3D shape descriptors. However, one assumption we have made is that local information and a small patch size in our learning is enough. Even though we already achieve good results, it would be worthwhile to explore higher resolution depth images and patch sizes to account for more global information, experiment with a larger number of 3D models, and incorporate convolutional methods to handle a larger network architecture.

Third, we can leverage information computed from existing shape analysis methods. For example, if we can first segment and label a 3D mesh [KHS10], we may be able to use this information to compute saliency values. A segmented "handle" already tells us that it is likely to be grasped. Another example is the information from assigning materials to 3D models [JTRS12]. A "softer" handle may be more likely and comfortable to grasp. Combining these ideas with our method can be a direction for future work.

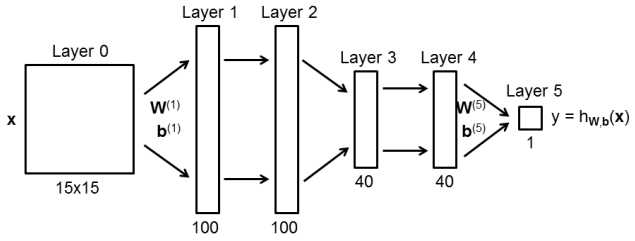

Figure 3: *Our deep neural network with 6 layers.* $\mathbf{x}$ *is a smaller and subsampled patch of a depth image and y is the patch center's saliency value. The size of each depth image is 300x300. We take smaller patches of size 75x75 which are then subsampled by 5 to get patches* ($\mathbf{x}$) *of size 15x15. This patch size corresponds to real-world sizes of about 4-50 cm. The number of nodes is indicated for each layer. The network is fully connected. For example,* $\mathbf{W}^{(1)}$ *has 100x225 values and* $\mathbf{b}^{(1)}$ *has 100x1 values. The network is only for each view or each depth image and we compute the saliency for multiple views and combine them to compute the saliency of each vertex. Note that we also need four copies of this network to compute the partial derivatives for the batch gradient descent.*

Acknowledgements

The original paper for this work was published at SIGGRAPH 2016 and we acknowledge our co-authors Weiqi Shi, Julie Dorsey, and Holly Rushmeier.

References

[BJB*15] BYLINSKII Z., JUDD T., BORJI A., ITTI L., DURAND F., OLIVA A., TORRALBA A.: MIT Saliency Benchmark. http://saliency.mit.edu/, 2015. 1

[GAGH14] GARCES E., AGARWALA A., GUTIERREZ D., HERTZMANN A.: A Similarity Measure for Illustration Style. *ACM Trans. Graph. 33*, 4 (July 2014), 93:1–93:9. 1

[IKN98] ITTI L., KOCH C., NIEBUR E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *PAMI 20*, 11 (1998), 1254–1259. 1

[JTRS12] JAIN A., THORMÄHLEN T., RITSCHEL T., SEIDEL H.-P.: Material Memex: Automatic Material Suggestions for 3D Objects. *ACM Trans. Graph. 31*, 6 (Nov. 2012), 143:1–143:8. 2

[KHS10] KALOGERAKIS E., HERTZMANN A., SINGH K.: Learning 3D Mesh Segmentation and Labeling. *ACM Trans. Graph. 29*, 4 (July 2010), 102:1–102:12. 2

[LDS*16] LAU M., DEV K., SHI W., DORSEY J., RUSHMEIER H.: Tactile Mesh Saliency. *ACM Trans. Graph. 35*, 4 (July 2016). 1

[LHLF15] LIU T., HERTZMANN A., LI W., FUNKHOUSER T.: Style Compatibility for 3D Furniture Models. *ACM Trans. Graph. 34*, 4 (July 2015), 85:1–85:9. 1

[LVJ05] LEE C. H., VARSHNEY A., JACOBS D. W.: Mesh Saliency. *ACM Trans. Graph. 24*, 3 (July 2005), 659–666. 1

[SEkB12] SAHBANI A., EL-KHOURY S., BIDAUD P.: An Overview of 3D Object Grasp Synthesis Algorithms. *Robotics and Autonomous Systems 60*, 3 (2012), 326–336. 1

[SMKF04] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The Princeton Shape Benchmark. *SMI* (2004), 167–178. 2

[VPV12] VARADARAJAN K., POTAPOVA E., VINCZE M.: Attention driven Grasping for Clearing a Heap of Objects. *IROS* (2012), 2035–2042. 1