Title:
Mosquito Popper: A Multiplayer Online Game for 3D Body Scan Data Segmentation

Authors:
Zachary Nolte, zachary-nolte@uiowa.edu, University of Iowa
Maxwell Riley, maxwell-riley@uiowa.edu, University of Iowa
Ramy Harik, harik@cec.sc.edu, University of South Carolina
Stephen Baek, stephen-baek@uiowa.edu, University of Iowa

Keywords:
Game with a purpose (GWAP); human computation; segmentation; labeling; human body scan.

DOI: 10.14733/cadconfP.2017.327-331

Introduction:
Game with a purpose (GWAP) is a concept that aims to utilize the hours spent in the world playing video games by everyday people to yield valuable data. The main objective of this research is to prove the feasibility of using the concept of GWAP for the segmentation and labeling of massive amount of 3D human body scan data. Accurate segmentation and labeling of 3D human body scan data is of paramount importance in many areas of research, such as human tracking [5], digital human modeling and anthropometric analysis [4],[12],[15], and etc. Such a task is extremely expensive as it requires lots of tedious, time-consuming hand labor. There are a number of computational methods for automatically segmenting and labeling a 3D model (e.g. [6],[8],[11],[13]), but the computers are still a way less reliable than the human capability in terms of accuracy, amount of detail, and robustness.

In an attempt to solve this problem and make geometry labeling and segmentation easier and less demanding, we created a multiplayer online game called "*Mosquito Popper*". The purpose of this game is to re-channel the efforts and the hours spent by game players to the collection of segmentation and labeling of through gameplay. Through crowdsourcing, it is possible to yield massive amounts of segmented mesh data while at the same time accurately label the 3D human body mesh segments. The advantage this approach offers is that it will help overcoming the problem of tedious and time-consuming work that previous computational approaches ran into. While playing the game, the players will effectively be providing geometry labeling and mesh segmentation data through the mechanics of the game. This makes the process of the tedious work and time spent more enjoyable and overall less laborious by utilizing the hours spent by people around the world playing games. Furthermore, by collecting this data directly from the players this way, we are using the natural 3D segmentation and image identification capability that humans already possess.

Related Works:
Game with a Purpose
In general, the main objective of GWAP is to accomplish a certain task by re-channeling the efforts devoted by game players. In particular, GWAP is beneficial for the tasks in which humans perform significantly better than computer algorithms, such as labeling, recognition, and segmentation. One of the early works on GWAP is the ESP game [1]. In this game, the players share an image, but this is the only common information they share. The goal of the game is for the players to come up with the same name or label for the image without communicating with each other. This concept in which randomly paired players try to achieve the same output is called the *output-agreement* mechanism. Through this game, it is possible to reliably label images and to help solve the problem of image recognition. As demonstrated by the ESP game, data accumulated through the time spent playing games can be used in a variety of practical applications.

Other examples of GWAP are Peekaboom [3] and Phetch [2]. Peekaboom locates individual objects contained within images and produces labels for them. One player reveals small pieces of a larger image to another player little by little. While the player revealing the image is trying to get the other player to guess a specific word, the smaller pieces are being labeled by the guessing player. Phetch annotates images with descriptive paragraphs with the intent of creating proper labels for images. Instead of relying on a web page author to create proper labels for each image on a website, the game has a player describe an image to a few other players, the seekers. The seeker's job is to web search for an image based on the description provided by the first player. The description provided by the describer is valuable data and becomes the text labels for the images being described. For more details, a comprehensive summary of previous work and games that have already been developed are contained in [10].

3D Model Segmentation and Labeling

Precise segmentation and labeling of a 3D model is one of the fundamental problems of computer-aided design. For the segmentation tasks, typically, one aims to design a computational algorithm that finds a perceptually sound partition of a geometry. For example, Liu *et al.* [11] mapped a surface geometry to the spectral domain and conducted k-means clustering to achieve the segmentation of a mesh model. Kalogerakis *et al.* [4] exploited the data-driven approach for the segmentation task. They formulated the task as a supervised learning problem in which they learned the objective function assessing the consistency between faces. Glovinskiy *et al.* [6] used a graph-based formulation to achieve consistent segmentation result across an entire set of models. They constructed a graph representing the similarity between faces, not just within the same mesh, but also across different meshes, and clustered the graph to achieve a consistent segmentation of multiple models of the same class. More recently, Harik *et al.* [7] defined a local shape signature based on the heat diffusion characteristics on a surface, and used the signature for assessing the likeliness of different faces belonging to the same segment. For more comprehensive review of a number of relevant works, see e.g., [13].

Although such computational approaches perform fairly well, they are still far behind the capability of humans. Even though a significant breakthrough has been made recently based on data-driven techniques, such as deep learning, and convolutional neural network [9], we still need an effective method for the segmentation and labeling tasks since those methods require a large volume of training examples as an input.

Method:

Game Play

In our work, a multiplayer game named *Mosquito Popper* was developed using the Unity 3D software engine, version 5.3.5f1 [14]. Using the terms of [10], the game mechanism we selected was the *problem inversion function computation*, in which a player must perform a computation based on some partial input to yield a result. Our mechanism in particular works by giving one of the two players access to all of the information while the other only has some. Based on this scheme, one player is assigned as a *describer*, and the other, a *guesser* in Mosquito Popper. Initially, a randomly-selected human model from a 3D whole body scan database is loaded in as well as a mosquito model. The human model is visible to the both players, but the mosquito model is rendered only on the screen of the describer. As the game begins, the mosquito randomly picks and flies to a vertex on the mesh of the human model starts sucking blood. The goal of the both players is to locate and smash the mosquito before it gets its fill and flies away. The guesser can smash the mosquito by mouse-clicking on the skin surface, but cannot see the mosquito and thus can only *guess* its location based on the information provided by the describer. On the other hand, the describer can see the mosquito on the screen but is not allowed to smash the mosquito. Therefore, the describer must communicate its location to the guesser through the use of a simple chat system implemented in the game. The describer may use proper and indicative words they see fit to describe the particular location. The guesser must click on the human model where they think the mosquito is based on the communication from the describer, ultimately trying to click on the mosquito and pop it. The guesser can click on the model as many times as they would like in order to find the mosquito. The both players earn points only if the guesser successfully "pops" the mosquito, and a new mosquito is spawned for the next game; however, the players achieve no points when the guesser fails to pop mosquito before the blood gauge displayed on the corner fills up and the mosquito flies away to a different location to start sucking blood again. Since the players share the common objective, they have a good motivation to cooperate well together. This stream of events continues until the main game clock runs out. A screenshot of gameplay can be seen in Fig 1. In Fig 1, the 3D human model is the white humanoid figure in the center of the image, the chat box is in the lower left corner of the image with the yellow text, the blood gauge and game timer are in the top right corner of the image. The rest of the scene, such as the checkerboard patterns, window art, and counters, is entirely cosmetic and is intended to provide a more fun, immersive experience for the players.

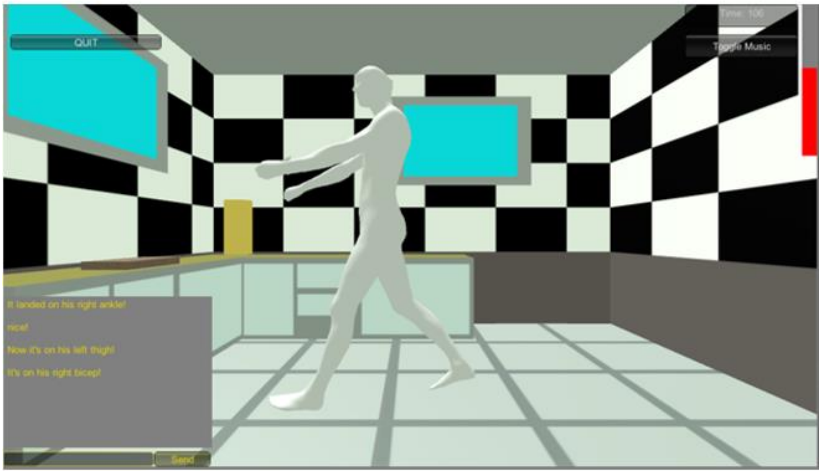

Fig 1: A screenshot of *Mosquito Popper* from the guesser's perspective.

Data Collection

While in the background, during gameplay, the game log data is extracted and collected. The information extracted from the game play is (a) the xyz coordinates of the points clicked by the guesser and (b) the chat elements provided by the describer. As the game is played, these data are added to arrays in real-time at the client computer. Once the game is completed, a word parsing algorithm is then run to extract the useful pieces of information out of the chat message log. This is performed by taking each element within the chat message array and comparing it to a *dictionary* of words pre-defined. This dictionary contains the words of interest, including adjectives/prefixes such as "fore-," "right," and "upper", and the names of the body parts such as "leg," "pelvis," and "arm." Therefore, if the describer communicated, "It's on the right upper arm!" the output from the algorithm would be, "right upper arm", eliminating the unnecessary and less useful pieces of information. The dictionary was created to include the vast majority of words that could be used to describe the human body, its function being something to compare the describer's input to, therefore allowing the users to communicate however they wished with no restrictions. The xyz coordinate values and the corresponding chat words are then sent to an online game server. The stream of information is then processed on the server-side and stored to a database. The results presented in this paper are, hence, created from the data collected on the server.

Segmentation and Labeling

The segmentation and labeling of the mesh occurs naturally through the gameplay of Mosquito Popper; the segmentation aspect takes place as the guesser is clicking on the mesh of the 3D human model and, the labeling aspect takes place as the describer is communicating the location of the mosquito. Computationally, each of the faces in the 3D mesh has bins for counting frequencies of vocabularies being mentioned by the describer, and the bins are updated according to the click history of the guesser. The similarity between the faces are calculated based on the bin counts in a similar sense to the well-known *bag of words* technique. Based on this metric, similar facets are clustered together to form a *segment*, and the bins are merged to determine the semantic label of the segment.

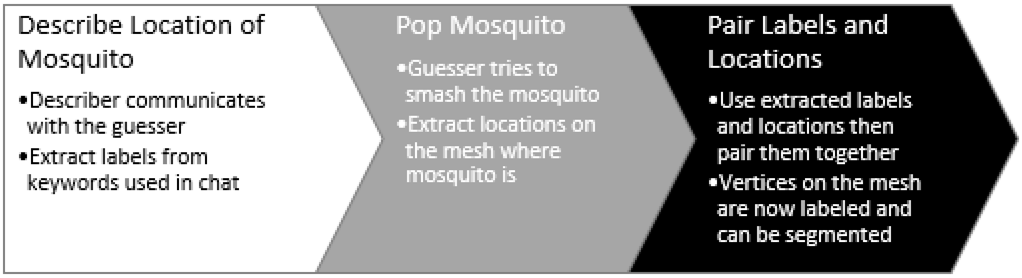

Fig. 2: A flowchart of the general process is shown. Each process element has a simple description of what is going on in the process and why it is important to the overall method.

<u>Results:</u>

In this section, we show the results of our experiments, the labels and segmentation acquired were generated automatically through gameplay. Our experiments included multiple trials using a single human 3D model. The trials were all conducted using the same group of 4 subjects. The only information the participants were given was the objective of the game. Each participant was urged to do their best to work with their partners to accomplish the task. Each pair of subjects played Mosquito Popper for an hour or two per trial, there were 6 trials performed by each pair. Based on the initial data we collected, some adjustments to the game were made. For example, we included more potential labels and added additional filtering to remove unwanted text from labels to increase utility and usability.

Results of our experiment is displayed in Fig. 3. Each of the dots annotates a clicked point on the surface by the smasher and the different colors represent different labels. Despite of the small number of trials during the experiment, the result still shows a promising outcome in terms of the accuracy and the level of details. We published our game on a website http://mosquitopopper.com and are collecting large amount of data from this online game. We expect to achieve high quality segmentation results for the human models in our database as we accumulate more hours of game play.

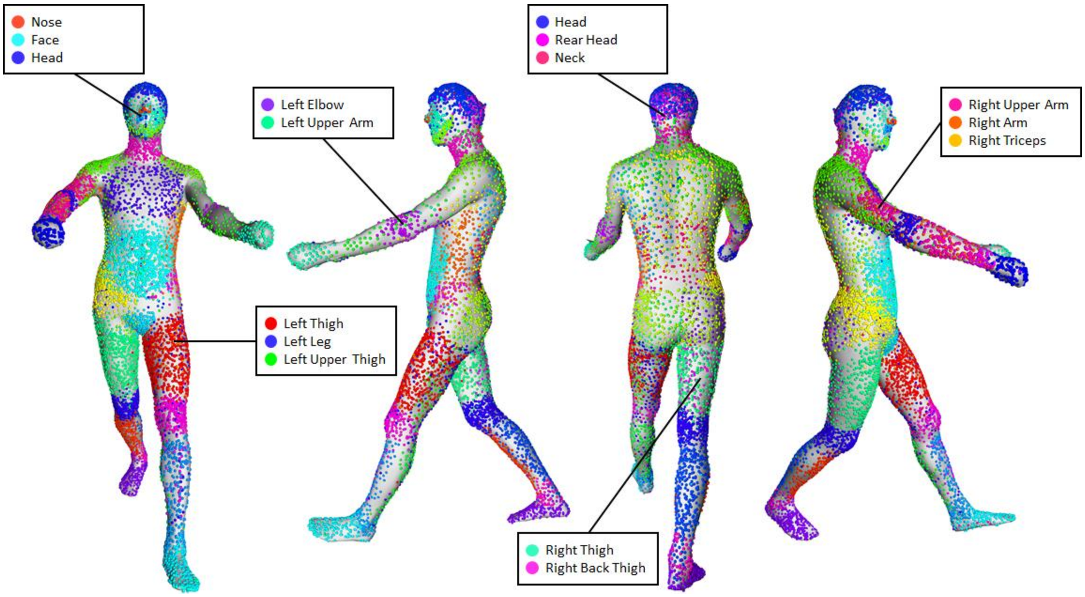

Fig. 3: Segmentation and labeling result from the experiment. Each of the dots on the human model represents a point of click by the smasher player, and the color denotes the label. Several examples of labeling per facet is also given in the text box.

Conclusion:

In this paper, we proposed a method of segmenting and labeling human body scan data with an online game. Our experiment and analysis showed that the proposed game play mechanism was effective for such a task, despite of a small number of hours of game play during the experiment. We expect a huge database will be generated in this way as the total hours of game play reach to a large enough level. The accumulated database can then be used as a reference and training dataset for various algorithms which will allow a computer to segment 3D human body shape data with minimal to no direct additional aid from a human user. This has huge implications as there are countless numbers of 3D models available on the web as well as 3D scan data being produced in labs and elsewhere.

Acknowledgements:

The authors thank Paul Yon and Timothy Dougherty for their technical assistance. The author would also like to thank all of the experiment participants that helped testing our game and producing our results.

References:

[1] Ahn, L.; Dabbish, L.: Labeling images with a computer game, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04), 2004, 319-326. https://doi.org/10.1145/985692.985733.

[2] Ahn, L.; Ginosar, S.; Kedia, M.; Liu, R.; Blum, M.: Improving accessibility of the web with a computer game, Proceedings of the SIGGCHI Conference on Human Factors in Computing Systems (CHI'06), 2006, 79-82. https://doi.org/10.1145/1124772.1124785.

[3] Ahn, L.; Liu, R.; Blum, M.: Peekaboom: a game for locating objects in images, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'06), 2006, 56-64. https://doi.org/10.1145/1124772.1124782.

[4] Baek, S.; Lee, K.: Parametric human body shape modeling framework for human-centered product design, Computer-Aided Design, 44(1), 2012, 56-67. https://doi.org/10.1016/j.cad.2010.12.006.

[5] Bălan, A.; Sigal, L.; Black, M.; Davis, J.; Haussecker, H.: Detailed human shape and pose from images, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, 1-8. https://doi.org/10.1109/CVPR.2007.383340.

[6] Golovinskiy A.; Funkhouser, T.: Consistent segmentation of 3D models, Computers and Graphics, 33(3), 2009, 262-269. https://doi.org/10.1016/j.cag.2009.03.010.

[7] Harik, R.; Shi, Y.; Baek, S.: Shape Terra: mechanical feature recognition based on a persistent heat signature, Computer-Aided Design & Applications, 14(2), 2017, 206-218. https://doi.org/10.1080/16864360.2016.1223433.

[8] Kalogerakis, E.; Hertzmann, A.; Singh, K.: Learning 3D mesh segmentation and labeling, ACM Trans. Graphics, 29(4), 2010, Article No. 102. https://doi.org/10.1145/1778765.1778839.

[9] Krizhevsky, A.; Sutskever, I.; Hinton, G.E.: ImageNet classification with deep convolutional neural networks, Proceedings of Advances in Neural Information Processing Systems, 2012.

[10] Law, E.; Ahn, L.: Synthesis Lectures on Artificial Intelligence & Machine Learning: Human Computation. Morgan & Claypool Publishers, San Rafael, CA, 2011.

[11] Liu, R.; Zhang, H.: Segmentation of 3D meshes through spectral clustering, Proceedings of the 12th Pacific Conference on Computer Graphics and Applications (PG'04), 2004, 290-305. https://doi.org/10.1109/PCCGA.2004.1348360.

[12] Reed, M.; Hu, J.: Rupp, J.: Statistical analysis of body shapes for parametric human body modeling, In Proceedings of the IRCOBI Conference on Impact Biomechanics Modelling, 2013.

[13] Theologou, P.; Pratikakis, I.; Theoharis, T.: A comprehensive overview of methodologies and performance evaluation frameworks in 3D mesh segmentation, Computer Vision and Image Understanding, 135, 2015, 49-82. https://doi:10.1016/j.cviu.2014.12.008.

[14] Unity 3D, https://unity3d.com/, Unity Software.

[15] Wang, C.: Parameterization and parametric design of mannequins, Computer-Aided Design, 37(1), 2005, 83-98. https://doi.org/10.1016/j.cad.2004.05.001.