A Comparative Study On Estimating Surface Orientation between Human and Neural Network

Ziwei Li, Renfei Wang, Shuhan Lu, Jiakun Gong, Avery Warner

Abstract— Human eyes have the most powerful ability to precept everything in our daily life. However, as the artificial intelligence has made significant progress in recent years especially in the area of computer vision, even machines are capable of perceiving images and learning from them. Thanks to the effective AI techniques such as neural network, traditional image tasks such as image classification and object detection can be handled by machines with high accuracy and performance. As such, in our study, we attempt to investigate whether AI can outperform human perception in estimating 3D surface orientations. Compared with previous image processing tasks, this is non-trivial because the perception on a 3D surface can be influenced by not only the terrain structure, but also the texture and shading. As a result, we also would like to study the accuracy of the surface orientations estimation with various texture and shading types. Therefore, In this study, we hypothesize that (1) adding texture on surface can increase estimation accuracy, (2) single light degrades the accuracy compared with combined lights, and (3) AI neural network outperforms human in terms of accuracy. To test on these hypothesis, we conduct an online human study by testing human-being's ability in estimating surface orientations. On the other hand, we build a neural network model to train such task with AI. Both studies are performed on a set of randomly generated 3D terrains with three types of textures and three types of lighting. Our experiment results reveal that AI neural networks can indeed achieve greater accuracy than human perception in such task. We also discover that the grid texture is able to reduce the estimation error, while combined lighting actually degrades the estimation accuracy.

Index Terms—Surface shape perception, textures, visualization, shape from texture, neural network.

1 Introduction

Human eyes are powerful, but what human eyes see in real life could still be inaccurate. In industrial production, there are cases when one needs to accurately know where a complex curved surface does, such as where human faces are facing when conducting medical screening. Such task may not be too difficult for humans to carry out, however, the evolution of the era of AI enables us to perform such tasks. With good amount of training, AI models seemed to be able to achieve perception tasks with high accuracy [3], which stipulates us to investigate the specific task of identifying surface direction, and answer is AI more accurate than human. In this paper, our goals are to: 1) determine whether textures would facilitate human to perceive the surface shape and surface orientation, 2) determine whether shading conditions would affect human to estimate the surface shape and surface orientation, 3) compare the overall performances between human and artificial neural network, and 4) determine whether textures would influence the accuracy of ANN under control of texture types. For our overall project anticipation, we are not only going to find out how textures and shading could influence our human perception of specific 3D surface and to find out how good could human see 3D objects under specific environment as well, but also we want to find out the artificial neural network could do the similar thing as for humans, and even if, Neural Network could do better than human. It will be a big challenge for us to find out the difference between human estimating result and AI estimating result. In this paper, we will create artificial surface and train both human and AI to estimate their surface normal. We will further conduct experiment on adding textures to the surfaces so that to see if that we make any difference in accuracy of estimation for both human and AI. Lastly, we will conclude features that could improve human or AI performance, so that to aid the progress in field of surface normal estimation. In our work, we have collected 14 samples from human studies designed by us and applied bootstrap to compare the efficiency of each type of texture as pairs. We did not apply the ANOVA which was used in [3], since essentially the data we collected will not satisfy normally collected data, which is fundamental assumption that needs to be fullfilled. ANOVA will draw inconclusive result with assumption violated. With bootstrapping, we are able to create replicate the samples and hence ease the problem of small sample. In our following discussion, we will discuss more about our set up of experiment on both human and AI and we will show that AI is indeed better at predicting orientation.

- *Ziwei Li is a student at the Ohio State University. E-mail: li.5326@osu.edu.*
- *Renfei Wang is a student at the Ohio State University. E-mail: wang.10876@buckeyemail.osu.edu.*
- *Shuhan Lu is a student at the Ohio State University. E-mail: lu.1743@osu.edu.*
- *Jiakun Gong is a student at the Ohio State University. E-mail: gong.204@buckeyemail.osu.edu.*
- *Avery Warner is a student at the Ohio State University. E-mail: warner.776@buckeyemail.osu.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

1.1 Surface shape perception

In field of shape perception, correct and effective representation of shape is one the most important topic of study as 3D objects were usually distorted for display and viewers would sometimes being trouble by not being able to interpret properties of 3D object such as where a specific point on the object is facing. To aid the representation of shapes, textures are commonly used to help viewers perceive the shapes. The use of textures had been researched in various studies [4], [6], it has been shown that by using different types of textures, the viewers accuracy of perception also varies with respect to textures. When trying to determine the orientation of shape in perception, according to [6], one needs to consider "1) Orientation of texture planes with respect to the viewpoint. 2) Orientation of texture planes with respect to the surface. 3) Orientation of viewpoint with respect to the surface. 4) Orientation of illumination with respect to surface/texture/viewpoint. Texture orientation, the illumination direction and the viewpoint may all interact in determining perceived shape." In our study, we are going to create our own textures for testing both human and AI's accuracy in predicting orientation of shape. We expect that using texture in shape perception could potentially aid the perception of shapes, especially in estimating orientation, so as to say, surface normal estimation. As a result, we carried out experiment focus on texture and lighting, in which we tested grid texture and curved texture, together with diffused lighting and ambient lighting. With the consideration of view point problem, we also implemented tool based on [6], which gives the human testers freedom to tilt and slant their predicted orientation to provide them more intuitive way of conducting experiment.

1.2 Surface normal prediction with neural networks

With evolved computing power and highly accessible tools developed, AI had achieved significant progress in computer vision. Neural networks are now able to predict image classification with great accuracy. With the capability of neural networks being proved to be true repeatedly, attentions had also been paid on using neural network to train specific tasks related to vision, for example, image depth estimation and normal estimation. For such tasks, simply reusing the existing deep neural networks will not satisfy the task, instead, more task specific architectures such as NYUDepth are developed for such purpose. For training neural networks that are used for surface normal estimation, one usually will take the input as the target image and use ground truth of each pictures for training the model. Originally, such surface normal prediction would be a regression problem in neural networks, however, with the surface normal triangular coding technique, the normal task of regression could be turned into classification task. As a result, one could render the input picture with visualization. Many complex models have been designed and tested for depth prediction and surface normal estimation. As in [3] and [6], the papers have shown that such neural networks model can even capture fine details in images. Such result showed us that neural networks has the potential in surpassing human for high accuracy prediction tasks for perception of direction in a finely detailed degree. We will present in the following section of more detail about surface normal estimation and carry out experiment on comparing human and neural network's performance.

2 Related work

In this section, we present a survey of related efforts in understanding human eye's perception system and AI perception system, and how different surfaces and textures are presented in visualizations and the performances.

There might require some prior knowledge which are critical for graph readings. Dubey et al. [2] gave us a glimpse of the human vision and the recognition system. They suggested that unlike computers, humans carry a great deal of prior knowledge about the world, which facilitates efficient decision making. This paper proposed that some general priors, such as the importance of objects and visual consistency, are critical for efficient game-play.

From the paper "View Direction, Surface Orientation and Texture Orientation for Perception of Surface Shape", Author Graeme et al. [6] found how textures, surface orientation and view direction are very important for perception of surface. They used two experiment of to see how textures and view angle will influence the result. The conclusions including textures constructed from planes more nearly orthogonal to the line sight and viewing it from oblique view is the best for perception. In another paper which has been posted on Current Biology Magazine [1], shows that mid-level vision is the one who did analysis and synthesis in our visual system. Author Barton L. Anderson shows that different shading and resurfacing will give us a different perception of the same object by mid-level vision. Three important features that mid-level vision does including extracting the sources of optical structure, shading the object, and finding the influence of specular reflectance. He shows an example of consequences of misattributing the cause of intensity gradients (shading and reflecting differences): two objects have different reflective properties but have same 3D shape. One is rendering matte, and the other will rendering as a combination of diffuse reflectance, and reflections could give us a more detailed shape of how this object looks. In "A Novel Cubic-Order Algorithm for Approximating Principal Direction Vectors", by Goldfeather et al. [5], they showed how tiny normal curvature approximation errors can be magnified into large errors in the estimated principal directions. Using a non-trivial test surface, they showed that principal direction approximation methods are easy to error when doing irregularly sampled data. However, the surface smoothing and noisy samples in meshes are further problems they need to fixed in further studies.

In the paper "designing Deep Networks for Surface Normal Estimation" [7], the paper is an early work on predicting or estimating surface normal of single image.The paper is one of the early works that tested the effectiveness of the use of deep neural network for sake of surface normal estimation.The method provided a top down, bottom up method that on the first top down phase, the deep neural network a structural estimation of surface normals for the image and a cuboidal approximation of the image through previous works' design of architecture. Then, in the second phase, the bottom up deep neural networks produce the surface normal and the edge label that distinguish objects and depths of the input image. The architecture produced accuracy close to the state of art architecture of the time on the task. This paper validated us the potential of using deep neural networks as a venue for estimating surface normal. In addition, many future works have shown the efficiency of using such a multi-stage method to estimate surface normal is an efficient method. For instance, in "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture" [3], the paper proposed a three staged convolutional neural networks that each of them are attributed to output features, predictions and refined predictions, and believe we can also attempt to implement a similar pipeline-like work flow.

In the paper "Visualizing intersecting surfaces with nested-surface techniques" [8], the paper is dedicated to create techniques that can merge and compare graphs together in one single picture instead of putting them side by side for comparison. The significance of this work is that the paper introduced the way one can compare the graphs, however, for our use, the paper helped us in giving hints on how to create the terrain we needed and merge it with texture to train different neural networks. In addition in paper "Textureshop: Texture Synthesis as a Photograph Editing Tool" [4], the author proposed a novel way of synthesizing textures for graphs without the need to reconstruct global 3-D mesh, and synthesize the texture instead on a network of individually parameterized surface patches. This paper is helpful in a way that it is directly related to our task of adding textures to images. It laid off both one way of theoretical foundation for us to understand the process and give us potential to look for tools that can help us implement textures into images.

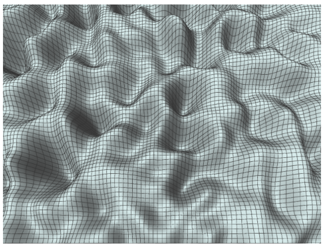
3 Hypotheses

- H1: Adding texture on terrain improves the accuracy of estimating the surface orientation; Grid texture outperforms the other two texturing conditions based on the results from previous literature [6].
- H2: Single lighting condition degrades the accuracy compared with combined lighting condition (i.e. combining ambient and diffused lighting together).
- H3: Artificial Neural Network outperforms the human subjects in estimating the surface orientation.

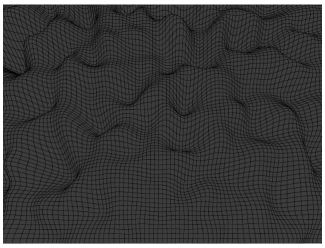
In H1 and H2, we conclude from related work that different textures and lighting will influence human's perception on shapes. Because neural networks were designed in a way that simulate human brain thinking, though not fully, still possess some characteristic of human thinking. We hence guess that textures and lighting may also have some impact on prediction accuracy of normal estimation. And for H3, our rationale is that, in recently days, we have seen numerous examples in AI outperforming human in repetitive tasks. For task of orientation estimation, we believe that AI may still have the advantage of outperforming human since if AI's learning won't be affected by the distortion of image, AI would be better at estimation. H3 indeed is trying to show that sometimes human may also be more biased than AI, as a result, one could explore more tasks to have AI to full fill in visual perception where humans may at disadvantage of viewing biased image.

4 Experiment I

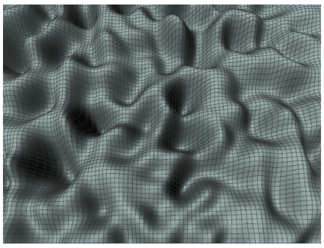
In our first experiment, we would like to conduct experiment to determine whether different lighting conditions will affect the accuracy for estimating the surface orientation. The task was to estimate the surface normal of a randomly generated terrain. Three different light



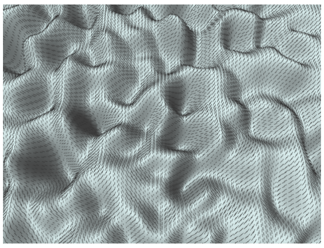
(a) Grid texture with combined lighting.



(b) Grid texture with ambient lighting.



(c) Grid texture with diffused lighting.



(d) CurveApp. with combined lighting.

Fig. 1: Combinations of textures with different lighting

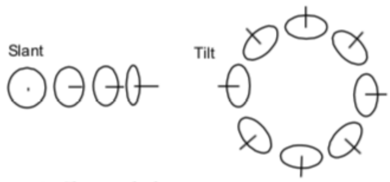


Fig. 2: Slant and tilt

conditions (i.e., ambient lighting, diffused lighting, and combined light) were examined in this study. We present a number of such terrain examples in Fig. 1, with different texture and light conditions.

4.1 Methodology

Graphical Design and Rendering Smooth random surfaces were constructed as our test terrains by summing 300 layers of Gabor function formulated as:

$$f(x,y) = k\cos\left(\frac{2\pi x}{w}\right)e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

The planes were set to orient at 68 degrees to the line of sight. Moreover, since two degrees of freedom are required to specify the orientation of a surface normal, we used a design of *slant and tilt* to represent the user input of surface orientation, as presented in Fig. 2. Slant refers to the angle between the normal of the surface and the line of sight; Therefore, it ranges from 0 to 90 degrees. whereas, tilt refers to the rotation of the cursor about the line of sight; Thus, it ranges from $+/-$ 180 degrees. In addition, to draw the test glyph, we disabled the Z-buffer test to resolve the occlusion issue caused from parts of the glyph.

Lighting Conditions Computer graphics illumination is used to simulate light sources and the way the light interacting with the objects in the scene. The implementation of the illumination techniques at a pixel level is called shading, which helps the 3D graphics reach a higher level of verisimilitude by demonstrating a three-dimensional shape. In addition, Both shading and the shape of the image contours generated by the bounding contours of shaded surfaces are linked to their local three-dimensional surface orientation [1]. Therefore, in this paper, we considered the influence of the existence of shading on estimating the surface orientation. To be more specific, we included two lighting components, which are ambient lighting and diffused lighting in the study to see if the lighting has any influence on estimating the surface orientation.

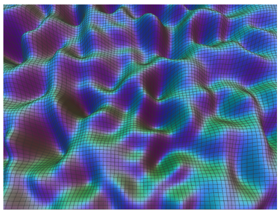


Fig. 3: Ground truth texture used for AI training.

Ambient lighting Ambient lighting is a general level of illumination and it is directionless. A typical example of the ambient lighting is shown in Fig. 1b. The amount of the ambient light that can be seen from a specific object is determined by the material of the object Ka and the ambient light contribution of each light source Ia. Both Ka and Ia are vectors for color R, G and B. The result of the ambient lighting will be the multiplication of these two vectors:

$$Ambient = Ka \times Ia \tag{2}$$

Diffused lighting As displayed in Fig. 1c, diffused lighting is the illumination that a surface receives from a light source and reflects equally in all direction. The amount of light the point on the surface receiving from a light source is determined by the Lambert's Law:

$$D = I \times \cos(\theta) \tag{3}$$

which I is the light intensity of the light source, theta is the angle between the vector from the point on the surface to the light source and the surface normal, and D is the amount of light the point on the surface receiving from a light source. Like the ambient lighting, diffused lighting is also determined by the material of the object Kd. Therefore, diffused lighting is calculated as:

$$Diffused = Kd \times Id \times \cos(\theta) \tag{4}$$

Combined lighting To combine those two lighting components together, we can simply add two lighting together, which results in the lighting condition shown in Fig. 1a. In our study, we only added one directional light so we only need to consider the ambient lighting and the diffused lighting based on one directional light. The shading model we used is Phong shading Model, which interpolates the normals and calculates the lighting for each pixel inside the polygon during rasterization.

4.2 Testing with human subjects

Subjects There were 14 subjects participated in our study, 10 undergraduate student and 4 graduate students. All experiments were conducted online and all participants were told that the accuracy was more important than the speed. However, even with tutorial, we still expect that test subjects will have difficulty carrying out the estimation since the fact that human eyes having trouble viewing distorted objects is an unmovable disadvantage.

4.3 Experimental Results

We used bootstrap approach to compute the mean and 95% confident interval of the experimental results, which shows in the Figure 5. We hypothesized that the combined lighting would provide better viewing environment and help to indicate fine details of the terrain surface. However, We observed that diffused lighting achieved the lowest absolute error compared with other two lighting conditions. Ambient lighting performed the worst among all the lighting types, which is not surprising giving that ambient lighting only represents the environment light.

5 Experiment II

The goal of this experiment is to determine whether textures have influences on the accuracy in estimating the surface orientation. Test Surfaces were generated with the same approach as illustrated in the first experiment. Whereas, different from our first experiment, artificial neural network is involved as another testing subject. Since we are interested in investigating how neural networks would perform in estimating the surface normal, especially compared with human perception under the same task.

5.1 Methodology

Under the same lighting condition, we examined three texture types in this study. We choose no texturing as our baseline of texture conditions, and the other two are the grid texture and the curvature approximating texture.

Grid Texture The grid texture combines both vertical and horizontal patterns together, which provides guideline for perceiving the surface orientation in two dimensions.

Curvature Approximating Texture Since the prior study has proven that the grid significantly outperformed the other texture types (e.g., contour, vertical, horizontal, and random), we would like to introduce another type of texture which indirectly provides the information of surface orientation. The proposed texture has a glyph shape, lying along both directions on the surface. Its direction changes depending on the curvedness of that region.

5.2 Testing with human subjects

Subjects Testing with human is expected to be a hard task, since determining certain orientations requires human eyes to be unbiased, which is rare. In our experiment, human test subjects need to be able to find the orientation of specific points on shapes. We have obtained 14 test subjects from universities, 10 undergraduate student and 4 graduate students. All experiments were conducted online and all participants were told that the accuracy was more important than the speed. However, even with tutorial, we still expect that test subjects will have difficulty carrying out the estimation since the fact that human eyes having trouble viewing distorted objects is an unmovable disadvantage. With such setup, human are tested on combinations of different lighting and texture or no texture, and result will be used for comparison with AI.

5.3 Experimental Results

Effect of texture to estimation accuracy We present the bootstrap results on three different types of textures (i.e., no texture, grid, and curve approximating texture) in Fig. 6. We find the mean absolute errors for the three textures are 0.984, 0.875, 0.953 respectively, and their confident intervals also appear to confront with this order, which indicates that the grid texture has the best performance followed by curvature approximating texture and no texture. Interestingly, the results prove with our hypothesis H1 that adding textures on the 3D terrain surface can help improve estimation accuracy. In particular, the grid texture has the best performance among the three textures, with 0.11 and 0.03 differences compared with the other textures.

Effect of lighting to estimation accuracy To study the effect of lighting to the estimation accuracy, we plot the bootstrap results of the three lighting conditions in Fig. 7. As shown, the diffused light has the best performance with a mean absolute error of 0.90, followed by combined light (0.94) and ambient light (0.96). Surprisingly, these results disprove our hypothesis H2 which assumes combined light condition can assist surface orientation estimation. We find that single light condition can actually achieve better performance, such as the diffused light compared with the combined light. We also observe that using the combined light is more like averaging two lighting conditions, since the absolute error of the combined light is roughly the average of the single lights.

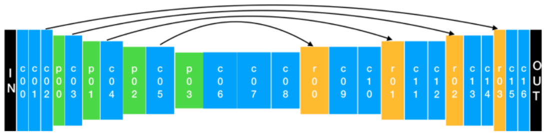


Fig. 4: The neural network architecture of our model.

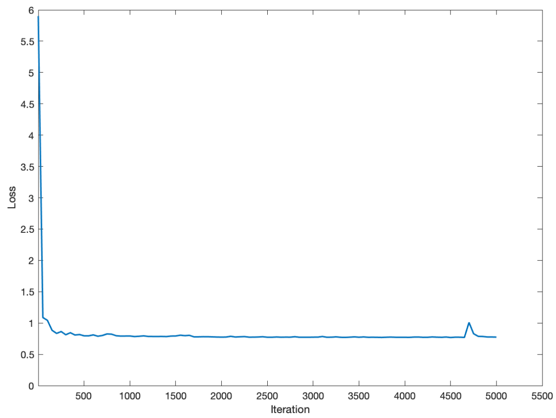


Fig. 5: Training results.

5.4 Testing with artificial neural network

Network Design and Experimental Setup We adopted a Hourglass Network as the base architecture, which combined inception modules and stack modules. The resulting triple-stack hourglass network is displayed in Fig. 4

Under each training procedure, the network takes three sets of input images, an original color image data set, a set of binary masks, and a set of color images mapped with ground truth normal values. The ground truth images were generated by following the algorithm for generating the normal map. A typical example of a normal map is shown in Fig. 3. Normal maps are commonly stored as regular RGB images where the RGB components correspond to the X, Y, and Z coordinates, respectively, of the surface normal. Each channel in the bitmap corresponds to a spatial dimension (e.g., X, Y and Z). An example ground truth image is shown in figure 3. Furthermore, Taylor expansion of "arccos" is used instead of $tf.acos()$ due to numerical instability, and the network was trained to minimize the total mean angular errors.

5.5 Experimental Results

We implemented the network algorithm in the Tensorflow and we used GPU device, Tesla K80, to do the network training. In each experiment, we generated 800 images for training and 200 images for testing, using three types of different textures (i.e.,, no texture, grid, and curve approximating texture). We plot the loss values for each training iteration in Fig. 5 using one running example. It can be inferred that the training converges quickly at around 200 iterations, where the loss value is around 0.8. We further train the model with 5000 iterations, and observe that the loss values do not have significant changes after 800 iterations, and the final training loss is around 0.77. In the following, we present the training and testing results in terms of three different textures.

The average absolute error of the no texture is approximately 0.315, and the average absolute error of the grid texture is approximately 0.281, while The average absolute error of the curve approximating texture is approximately 0.372. It is obvious that the grid texture has the best estimation accuracy, followed by no texture and finally the curve approximating texture. The order confronts with our human study results. As for the overall accuracy, the AI training can achieve significantly higher accuracy than human beings (by around 0.6 absolute error difference), which proves our hypothesis H3. Moreover, we do

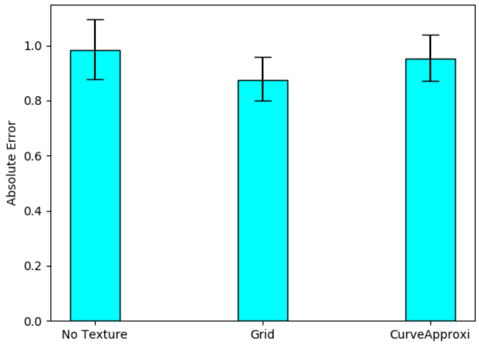


Fig. 6: Bootstrap results by texture type (human study).

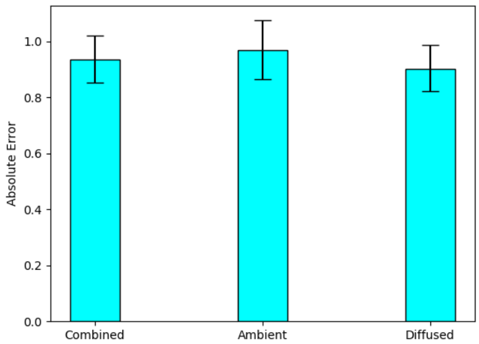


Fig. 7: Bootstrap results by shading type (human study).

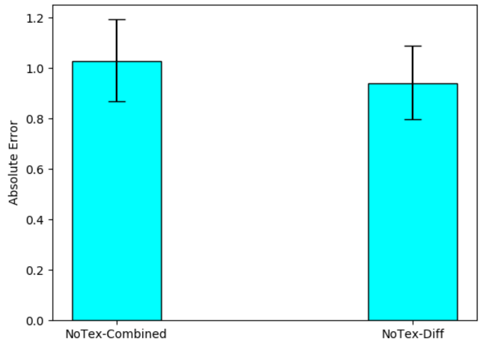


Fig. 8: Bootstrap results by no texture (human study).

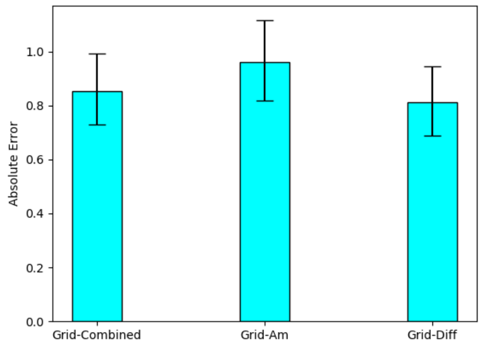


Fig. 9: Bootstrap results by grid (human study).

not observe significant difference between these three textures in AI training. We suspect that this is due to the low resolution of images.

6 Discussion

Since terrain shape influence the results for both human and network, however, there are still some differences in between them. We did both Human testing and Neural Network training for AI, the results shows that AI has relative higher accuracy in estimating the surface orientation. This does not surprise us, because we anticipate that AI could do better than human.

When doing the human experiment, we did a recall of required knowledge and teach them how to use glyph before our experiment. Required knowledge for understanding how this experiment working includes but not limited to: Normal Value, Ground Truth, 3D vision. We also did a short survey asking their feelings about this experiment after finish the whole test. Many of our test takers do not need the recall for required knowledge, however, they still get lost once they are in the experiment. They frequently asked questions what they need to do and how glyph works during the test. This is sign that the introduction and basic manual for our experiment is confusing and not really complete, and this is the point we need to fix. The comments from participants after our experiment varies. Some of them think the test is too long and hurt their eyes. Most of them give feedback of some shading and textures are hard to estimate. Textures they pointed out that hard to see are no texture and curvature approximating texture. Shadings they are hard to use them in estimation are no lighting and only ambient lighting. no texture with no shading is the hardest one to do. Combined our result, we find that ambient lighting performed the worst among all the lighting type, and grid texture performed the best among all texture type, which match the comments from test takers.

Terrain we are currently using is kind of uneven, which means it does not have a very smooth surface. We created a terrain for NN training before has a smoother surface, which could achieve higher accuracy in network testing. However, the terrain we use for human experiment is not smooth, but it is easier for human to estimate the normal values and surface orientation for this newer terrain.

In human testing, view angle from terrain also influence the accuracy for network. Since human only have a fixed angle to see the 3D model, which means if a point unfortunately selected on the back side of a slope, and that is the blind spot for human, and test takers will just simply believe that it is on the front slope and can be estimated. This will increase the error a lot, cause the opposite position for that point might have totally different slant and tilt.

Another problem for our AI training is the image resolution. Since the terrain used to be big, but fitting inside our training model, we need to resize the images which will degrade the quality of images. That will actually increase the chance for accidentally wrong estimation in Neural Network. We did not have any comments from human test takers reporting size of image will cause problem for their estimation.

7 Conclusions and Future Work

From the experiments conducted, the results show that neural networks perform better on estimating the surface orientation. The Bootstrap result shows that, for human participants, both textures reduce the surface orientation estimating error and variance, especially for the grid texture. Also for human participants, the diffused lighting reduced the surface orientation estimating error and variance. Our experiment result shows the similar result as our hypothesis. From the comments of participants, they stated that it is hard to estimate under no texture and only with ambient lighting, which also matches our experiment result. In the future, we may further improve the neural networks to make it work better with the rugged and tilted surface. We may also

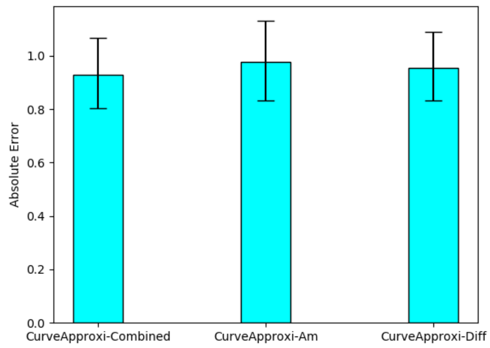


Fig. 10: Bootstrap results by curveApproxi (human study).

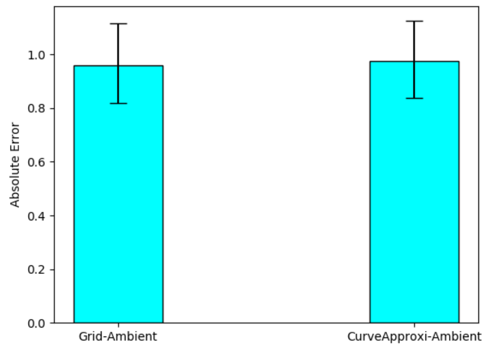


Fig. 12: Bootstrap results by ambient lights (human study).

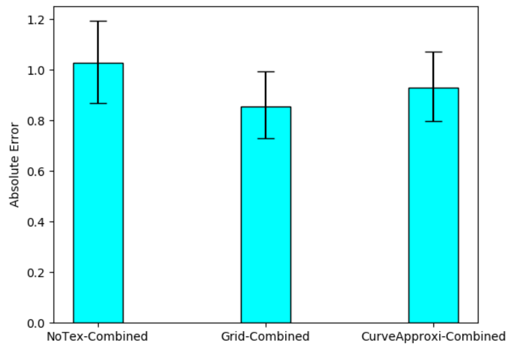


Fig. 11: Bootstrap results by combined lights (human study).

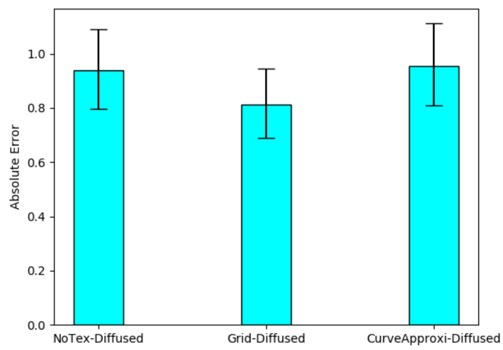


Fig. 13: Bootstrap results by diffused lights (human study).

add color as a visual cue (e.g. colormap) in the task to see if it improves the surface orientation estimating result. Finally, the image resolution is not perfect for now and we may experiment it with images with a higher resolution. For future work, we recommend further studies to first collect more samples for human test, and train people with more related specialization to test the experiment and compare with neural networks. We also recommend optimizing the use of more powerful and state-of-art AI models with more efficient architecture, so that one can really obtain the specialized result in terms of the normal estimation. Further, we also propose that understanding why adding textures will ease perceptions for AI and humans and it will be interesting to see the features that explains pertained to textures that make prediction better. In recent studies, various tools had been developed to explain how AI come up to their specific conclusion, we believe such tool could help future studies to conduct why adding various features from textures will change the prediction accuracy. In our work, we tended to verify our assumptions through experiment, we believe that future work can attempt to develop more concrete reasoning on normal estimation and why textures help.

References

[1] B. L. Anderson. Mid-level vision. *Current Biology*, 30(3):R105–R109, 2020.

[2] R. Dubey, P. Agrawal, D. Pathak, T. L. Griffiths, and A. A. Efros. Investigating human priors for playing video games. *arXiv preprint arXiv:1802.10217*, 2018.

[3] D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE international conference on computer vision*, pp. 2650–2658, 2015.

[4] H. Fang and J. C. Hart. Textureshop: texture synthesis as a photograph editing tool. *ACM Transactions on Graphics (TOG)*, 23(3):354–359, 2004.

[5] J. Goldfeather and V. Interrante. A novel cubic-order algorithm for approximating principal direction vectors. *ACM Transactions on Graphics (TOG)*, 23(1):45–63, 2004.

[6] G. Sweet and C. Ware. View direction, surface orientation and texture orientation for perception of surface shape. In *Proceedings of Graphics Interface 2004*, pp. 97–106. Canadian Human-Computer Communications Society, 2004.

[7] X. Wang, D. Fouhey, and A. Gupta. Designing deep networks for surface normal estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 539–547, 2015.

[8] C. Weigle and R. M. Taylor. Visualizing intersecting surfaces with nested-surface techniques. In *VIS 05. IEEE Visualization, 2005.*, pp. 503–510. IEEE, 2005.