PERCEPTUAL EVALUATION AND METRIC FOR TERRAIN MODELS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Suren Deepak Rajasekaran

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Bedrich Benes, Chair
Department of Computer Graphics Technology
Dr. Martin Cadik
FIT, Brno University of Technology; FEL, Czech Technical University
Dr. Tim McGraw
Department of Computer Graphics Technology
Dr. James L. Mohler
Department of Computer Graphics Technology

Approved by:
Dr. Kathryne Newton
Head of the Graduate Program

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Bedrich Benes, for guiding and motivating me throughout this research and graduate school. My committee members, Dr. Martin Čadík, Dr. Tim McGraw and Dr. James Mohler for their inputs and unique perspectives on this research topic.

Furthermore, I would like to thank Hao Kang, a fellow HPCG Lab member and friend, without whom this dissertation topic would not have come to fruition.

This research work is a product of collaboration with Dr. Bedrich Benes (Purdue University, USA), Hao Kang (Purdue University, USA), Dr. Martin Čadík (FIT, Brno University of Technology; FEL, Czech Technical University in Prague, Czech Republic), Dr. Eric Galin (Université Claude Bernard Lyon 1, France), Dr. Eric Guérin (Université Claude Bernard Lyon 1, France), Adrien Peytavie (Université Claude Bernard Lyon 1, France) and Dr. Pavel Slavík (Czech Technical University in Prague, Czech Republic). It has been submitted to SIGGRAPH ASIA 2019: The 12th ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia.

Additionally, I would also like to thank Terragen and Vue (e-on Software) for providing student license for their software packages. This research was funded partially by National Science Foundation grants #10001387, *Functional Proceduralization of 3D Geometric Models*.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

ABBREVIATIONS

2D	Two-Dimensional
3D	Three-Dimensional
CG	Computer Graphics
CGAN	Conditional Generative Adversarial Neural Network
CGT	Computer Graphics Technology
GAN	Generative Adversarial Neural Network
HVS	Human Visual System
IPM	Inverse Procedural Modeling
IRL	Inverse Reinforcement Learning
ML	Machine Learning
PM	Procedural Modeling
PTQM	Perceived Terrain Quality Metric

ABSTRACT

Rajasekaran, Suren Deepak Ph.D., Purdue University, August 2019. Perceptual Evaluation and Metric for Terrain Models. Major Professor: Bedrich Benes.

The use of Procedural Modeling for the creation of 3D models such as Buildings, Terrains, Trees etc., is becoming increasingly common in Films, Video Games, Urban Modeling and Architectural Visualization. This is due to the primary factor that using procedural models in comparison to traditional hand-modeled models helps in saving time, cost and aids in generation of a larger variety in comparison to a few. However, there are so many open problems in procedural modeling methods that does not rely on any user assistance or aid in generating models especially in terms of their visual quality and perception. Although, it is easy to identify realistic looking models from procedural models, the metrics that make them 'Real' or 'Procedural' is still in the indeterminable and remains uncanny in nature. The perceptual metrics (intrinsic factors such as surface features and details, extrinsic factors such as environmental attributes and visual cues) that contributes to the visual perception of Procedural models have not been studied in detail or quantified yet.

This dissertation presents a first step in the direction of perceptual evaluation of procedural models of terrains. We gathered and categorized several types of real and synthetic terrains generated by methods used in computer graphics and conducted two large studies with 70 participants ranking them perceptually. The results show that synthetic terrains lack in visual quality and are perceived worse than real terrains with statistical significance. We performed a quantitative study by using localized geomorphology based landform features on terrains (geomorphons) that indicate that valleys, ridges, and hollows have significant perceptual importance. We then used generative deep generative neural network to transfer the features from real terrains to synthetic ones and vice versa to further confirm their importance. A second perceptual experiment with 128 participants confirmed the importance of the transferred features for visual perception. Based on these results, we introduce PTQM (Perceived Terrain Quality Metrics); a novel perceptual metrics based on geomorphons that assigns a number of estimated visual quality of a terrain represented as a digital elevation map. The introduced perceptual metric based on geomorphons indicate that features such as Valley (0.66), Ridge (0.64), Summit (0.44), Depression (0.42), Spur(0.33), and Hollow (0.22) in order have significant perceptual importance. By using linear regression, we show that the presented features are strongly correlated with perceived visual quality.

CHAPTER 1. INTRODUCTION

Modeling is a fundamental component in the asset creation pipeline of many areas in computer graphics. 3D models of any kind or variety are essential in Architectural Visualization, Urban Modeling, Films, Games and many other areas that utilize them. These assets may primarily be the object of focus in the context of a given scene or simply be a part of it. Modelers who primarily work on creating these assets dedicate considerable amount of time and effort in sculpting these models from scratch. This process is usually referred to as Hand Modeling or Traditional Modeling approach. Therefore, there has always been research in the modeling field for easing the process of asset creation. As the field of computer graphics became increasingly complex and sophisticated, so are the methods for synthesizing models such as Hand Modeling techniques, Data-Driven Modeling techniques (use of data in aiding the modeling or the design process) and procedural methods that use mathematical methods with formal grammers for representing and generating 3D models and their textures.

Terrains are among the most visually stunning structures in landscapes and their modeling has attracted attention of computer graphics researchers for decades. Patterns found in terrains result from eons of complex and interacting geomorphological processes with varying strength at differing spatial and temporal scales, which makes them extremely difficult to capture. Additionally, other phenomena, such as the underlying bedrock strata, tectonics forces, vegetation development, temperature influence, or presence of glaciers, make terrain modeling even more complicated.

Humans experience terrains through their entire life and our visual perception system has evolved into a very precise tool for judging the overall appearance of terrains. Humans are excellent in detecting minor anomalies in them

Travers (1984), which makes synthetic terrain modeling the more challenging as quantifying those inconsistencies remains highly complex. Generating geometric models of naturally looking terrains is a difficult task that usually requires expertise. Although a wide variety of algorithms exists for modeling terrains (see the recent review Galin et al. (2019)), existing methods consider the morphological phenomena shaping the terrains independently and their mutual dependencies are neither well-studied nor understood.

Procedural modeling methods in the computer graphics field is very vital and has been widely utilized in films and game production areas (but not only limited) to save time. However, there are three major modeling fields where procedural modeling tools are being used extensively for model generation: buildings, trees and terrains. This is highly evident from the sales and prominence of procedural modeling oriented toolkits such as Houdini (SideFX (2019)), SpeedTree (SpeedTree (2019)), Terragen (PlanetsideSoftware (2019)), Vue (E-onSoftware (2019)), XFrog (XfrogInc (2019)) etc., for use in these aforementioned areas. Procedural Modeling is the process of using generative methods to create a 3-D model of effect (Smelik, De Kraker, Tutenel, Bidarra, and Groenewegen (2009) & Smelik, Tutenel, Bidarra, and Benes (2014)). Procedural modeling is an alternative to manual modeling and it includes advantages such as data compression, time-saving attributes, and variety Smelik et al. (2014). Procedural model generation methods can easily be able to produce a variety of models with a quick turn around time from a single representation. These attributes made procedural modeling to be widely adapted in wide areas of application. There are many different methods of procedurally generating terrains can be classified into two main categories: 1) stochastic methods and 2) erosion or growth models and weathering simulations. Procedural modeling methods albeit having a lot of advantages tend to have an important drawback when compared to other modeling methods which is the lack of control over generated models. Though there are significant advances in this field, there exists a research gap in which the generated models are not evaluated in terms of how they are perceived by the audience that it is intended for. This gap has been addressed in procedural texturing area where the results show that the textures are not perceptually meaningful as they tend to be ”repetitive,” ”directional,” ”structured,” (Liu, Dong, Cai, Qi, and Chantler (2015a)). This requires the need for studying the characteristics or procedurally generated models as to understand their visual perception metrics so that purely procedural methods can be sufficient in terms of satisfying the model requirement needs with out the need of any user assistance.

While the terrain morphology is an important driving force behind the landscape formation, the human observer is the final judge of the computer generated model's visual plausibility. Previous methods focused on replicating phenomenological processes of terrain formation, but none, to the best of our knowledge, have focused on the human perception of terrain models. The evaluation of results of algorithms simulating natural phenomena has been a tedious question for almost everyone working in this field and is usually addressed by providing side-by-side comparison of the generated structures or is assumed to be correct if the underlying simulations are physically-based.

This dissertation is a first step in the direction of rigorous perceptual validation of computer graphics generated models of terrains. In particular, we attempt to answer the question of what are the visually important features in terrains that make them visually plausible, and what is the level of visual plausibility of synthetic terrains commonly used in computer graphics. A recent work in geology allows for quantitative evaluation of terrain by using so called geomorphons that are set of quantitative features assigned to digital elevation maps Jasiewicz and Stepinski (2013). We performed an extensive user study, in which we measured the perceived visual plausibility of real and synthetic terrains. We express the plausibility in terms of the underlying geomorphons. We then used the state of the art deep neural networks CycleGAN Zhu, Park, Isola, and Efros (2017) to transfer features from the images that were ranked high to those ranked low and vice versa. We performed another user study that shows that the landforms transferred from highly ranked sets to lowly ranked ones improve the visual perception and that the landforms transferred from low-ranked images to high ranked ones demote them perceptually. Results of the two user-studies combined with the analysis of features show that synthetic terrains do not often include commonly found geomorphological features such as *depressions*, *summit*, *flat*, *valley*, *ridge*, *hollow* and *spur* geomorphons. We introduce PTQM (Perceived Terrain Quality Metrics) that assigns a normalized value of perceived perception to a terrain represented as a digital elevation model.

The primary motivation in attempting this research is in identifying the perceptual attributes of procedurally generated natural objects. We hypothesize that the identified visual perception metrics can be ranked and used in improving the existing procedural modeling methods which are very vital in thoroughly automating the modeling process. We begin this study with the assumption that those perceptual attributes can be identified and measured with the use of a mixed methods approach and the same can be used in machine learning approaches to classify perceptually realistic models and unrealistic models.

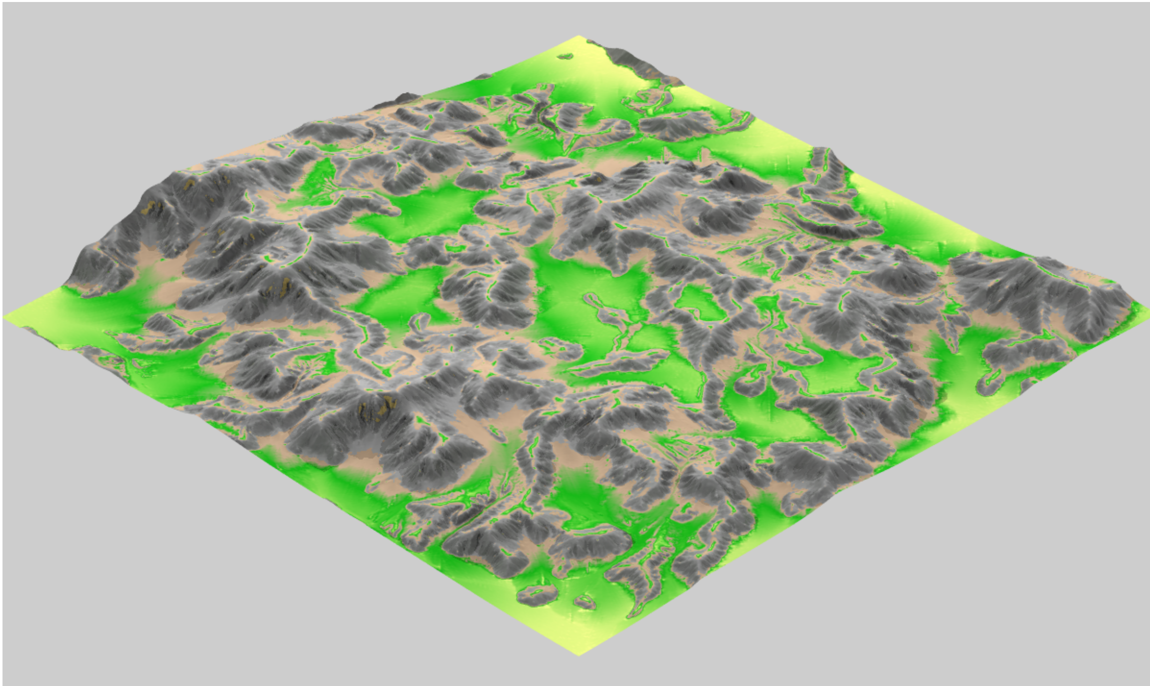


Figure 1.1. Rendered example of a procedural terrain digital elevation map generated with fluvial erosion patterns.

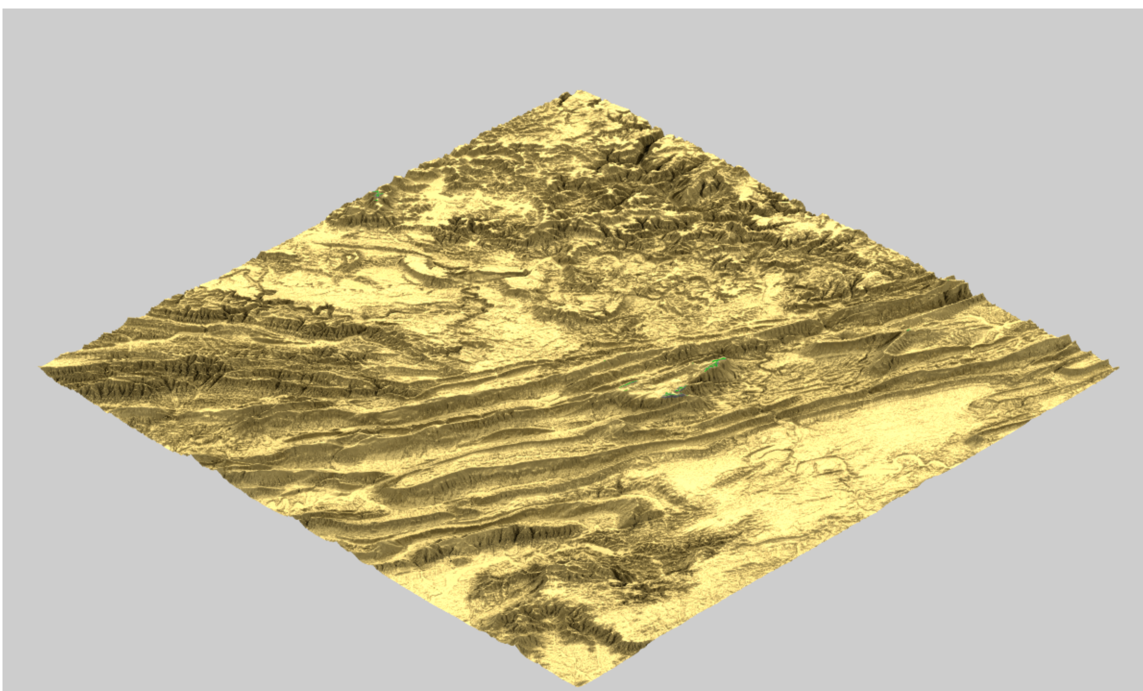


Figure 1.2. Rendered example of a real terrain digital elevation map with fluvial and aeolian erosion patterns from South Carolina state in USA.

1.1 Scope

The scope of the research is to identify and list the metrics that directly affect the visual perception of Terrain models. Our goal is to improve the purely procedural modeling methods by using these metrics and automate the process even further without any interventions from the user for producing models that are visually plausible. Therefore, we attempt to train a neural network to identify and improve visually important features that may be imperceptible to the human eye.

1.2 Significance

There are three major contributions of this dissertation work:

1. Identification of visual perception metrics in the context of procedural models of terrains.
2. Improving the quality of terrain models that are generated based on purely procedural methods without user assistance.
3. Additionally, the methods shown in this study are scalable and reapplied to other categories of procedural generation methods.

1.3 Research Question

1. Are procedurally modeled terrains visually perceived by humans equally as the real terrains ?
2. What are the elements of terrains that have the strongest effect on perceived quality ?
3. Is it possible to change the perceived quality by transferring elements from real and procedural terrains by the use of a Neural Network ?

1.4 Assumptions

The assumptions for this study include:

- It is possible to identify and quantify the perceptual factors that affect the perception of procedural Terrain models through a mixed-methods approach.
- The identified subset of procedural features that are visually important can be parameterized and mapped to the inputs of procedural models.
- The visual features of a Terrain model can be perceptually improved by the utilization of a Neural Network.

1.5 Limitations

The limitations for this study include:

- The research work targets only the intrinsic factors that affect the perception of procedural models and a detailed study of extrinsic perceptual factors is beyond the scope of the study.
- The work is also limited only to the perception of shape and physical features of the procedural models rather than the effect of rendering modes or textures in their representation.
- The training data was not readily available and had to gathered, prepared, annotated from scratch, even though the user studies are carefully designed, this presents us with a challenge of limited validity of results.

1.6 Delimitations

The delimitations for this study include:

- Our method only focuses on attempting to improve the procedural model generation methods that are purely automatic (such as Noise-based generation methods for Terrains) and not thoroughly user assisted to represent their real world counterparts.

- Trained and Expert users of Terrain models such as Geologists are not considered in terms of measuring the accurateness of visual perception.
- The participants in the user study may introduce personal perceptual bias in the survey and interview procedures that is beyond the control of researchers.
- The developed technique is primarily dependent on the data used for training the neural network therefore there are no guarantees about generalization of the technique.

1.7 Definition of Key Terminologies

In the broader context of thesis writing, we define the following terms:

Framework: "Frameworks model a specific domain or an important aspect thereof. They represent the domain as an abstract design, consisting of abstract classes (or interfaces). The abstract design is more than a set of classes, because it defines how instances of the classes are allowed to collaborate with each other at runtime. Effectively, it acts as a skeleton, or a scaffolding, that determines how framework objects relate to each other." (Riehle, 2000).

Inverse Procedural Modeling: The inverse process of a Procedural Modeling which involves estimating the formal grammer and parameters of a given model.

Machine Learning: Machine Learning is a field of Aritificial Intelligence that uses statistics and data to drive learning in machines without the need for explicit programming.

Mixed Methods Study: Mixed methods study is a type of research that involves the use of more than one method to aid the data collection process.

Neural Network: Neural network is a machine representation of biological animal brain that aids the process of machine learning in computers.

Perception: Perception is the interpretation of sensory information received by the brain in order to understand.

Procedural Modeling: Procedural Modeling is the process of using generative methods through the used of Mathematical models to create 3-D model (Smelik et al. (2009) & Smelik et al. (2014)).

Thematic Analysis: Thematic analysis is a data analysis technique used in Qualitative research studies to identify the underlying "themes" and "patterns" in a data set.

1.8 Summary

This chapter provided the scope, significance, research question, assumptions, limitations, delimitations, definitions, and other background information for the research project. The next chapter provides a review of the relevant literature that will cover existing procedural modeling techniques, perception issues, mixed methodology research, machine learning and recent developments in the same fields.

CHAPTER 2. REVIEW OF RELEVANT LITERATURE

This literature review attempts to showcase important information relevant to the areas and research questions that the work is trying to address. This section is divided into multiple sections: First, the Human Visual System (HVS) and how perception works is briefly showcased followed by general perceptual issues in Graphics. Second, Perceptual metrics and how perception driven graphics approaches have aided in improving computer graphics methods has been studied. Finally, specific perceptual issues in Procedural Modeling are discussed, followed by general factors that are influential in perception of a specific category of model type.

2.1 Visual Sensitivity from HVS & Psychology of Perception

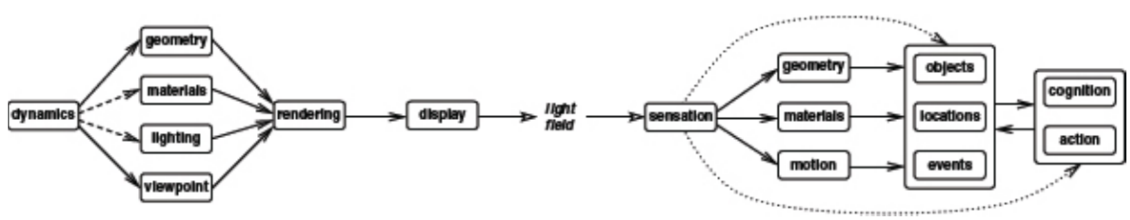


Figure 2.1. Description of World from a Human Eye (Thompson et al., 2011)

Computer Graphics, principally, is a field that is 'visual' oriented, the main purpose of the area is to synthesize images for people to view (Thompson et al., 2011). The HVS that consists of the eyes and complex neural network, becomes the epicenter of everything that is synthesized and simulated graphically for viewing. The above included graphics exemplifies how a dynamic scene is broken down visually by the human visual system. The understanding and study of human visual system is in the amalgamation of fields such as perceptual psychology, eye-movements, medicine, physics, neuro-science and computational vision (OSullivan, Howlett, Morvan, McDonnell, & OConor, 2004). The concept of perception is subjective, given how complicated the Human visual system is, 'overgeneralizing' in this field usually ends up giving shoddy results. Therefore, there is a need for perceptual effectiveness in the research that is conducted, so that the computer graphics imagery, "looks" as they are intended to be (Thompson et al., 2011).

The understanding of how a computer generated object as perceived by human eye, is core to the overall scheme of things because it helps in maximum throughput while being efficient by not focusing on the things that are not essential to the HVS (Bartz, Cunningham, Fischer, & Wallraven, 2008). Fundamentally, all the Image, visual and perceptual quality metrics and enhancement approaches is based upon the HVS ((Saghri, Cheatham, & Habibi, 1989) & (Panetta, Wharton, & Agaian, 2008)).

2.2 General Perceptual Issues and Factors in areas of Computer Graphics

Perception oriented studies has been in existence for a long time, not necessarily in CG, but in the fields of Psychology, Art, Visualization. One of the pioneering research in the field was by Gibson (1950) for studying the perception of textures. However, in CG, albeit not as broad as psychology, perception has been studied in all subfields such as how distances, synthetic human models, animations, emotions, scale, depth, and effects of motion are studied in both real and virtual environments (Virtual Reality and Augmented Reality) ((Loomis & Knapp, 2003) & (Jones, Swan II, Singh, Kolstad, & Ellis, 2008)). Perceptional issues specific to Augmented Reality area in Computer Graphics has been studied in detail a couple of times by Drascic and Milgram in 1996 and Kruijff, Swan, and Feiner in 2010. Hodgins, Jörg, O'Sullivan, Park, and Mahler studied how the human anomalies in simulated human characters are easily identifies by the audience based on the famous 'Uncanny Valley' response in the year 2010. A thorough detail of human visual system, perceptually motivated rendering, optimizations, and how perception drives computer graphics combined with new trends in perception oriented graphics research for image processing, facial animation, realistic characters in movies has been presented by McNamara, Mania, and Gutierrez (2011) as a SIGGRAPH course in the year 2011.

2.3 Perceptual Metrics in Computer Graphics

The various subjective perceptual phenomenon of the HVS has been quantified by researchers several times as a metric that are highly useful for measuring and understanding it. This is evident from the studies that attempted in measuring a subjective HVS phenomenon such as 'Just-Noticeable-Difference' (JND) in visual quality for 3D models by reducing the number of polygons and generating a threshold parameter (Cheng, Firouzmanesh, & Basu, 2011). A detailed research of measurement metrics for perceptual visual quality metrics in Images has been studied and listed by ((Gao, Brooks, & Arnold, 2017) & Lin and Kuo (2011)). There has been numerous researches and comparative studies on the perceptual metrics of 3D models such as model Quality, for both static and dynamic triangular models ((Cleju & Saupe, 2006), (Lavoué, 2011), (Corsini et al., 2013), & (Elloumi, Kacem, Dey, Ashour, & Bouhlel, 2017)). On the same lines, a perceptual metric study that encompassed 3D models along with Light and Material Interaction had been conducted to show how the model modification impacts on their perception (Vanhoey, Sauvage, Kraemer, & Lavoué, 2017).

Just-Noticeable-Difference is another common and useful perception metric in Computer Graphics and has been useful in measuring the degree of perceptual change in areas such as Animation, View Estimation, Texture and Material perception etc., Just-Noticeable-Difference metric for models (static 3D objects) has been devised by (Cheng & Boulanger, 2005) to analyze different types of models to identify redundant information so that available bandwidth can be used for improving texture resolution without compromise in perceived quality. Cheng and Boulanger's method focuses on viewing distance independent static models and also accounts for changes in varying textures and lighting. The findings of this paper proves that not every set of vertices in a 3D Model has 'significant impact on visual quality' and each model has its own unique surface property, perceptual values. A drawback of this study is that only one type of model (Ellipsoid) is analyzed and a certain portion of (median axis of the model) of the 3D Model; It is highly dependent upon the model refinement approach that is being used, therefore the results that are from this experiment are not very generalizable to organic model varieties such as a Terrains or Trees.

2.4 Perception-Driven Graphics Approaches

The knowledge of human perception has been applied in computer graphics to optimize the modeling and rendering pipeline since the beginnings and a common way is to incorporate it as a computational model of a particular HVS feature, visual masking (Ferwerda, Shirley, Pattanaik, & Greenberg, 1997), visual attention and saliency (Frintrop, Werner, & Garca, 2015; Riche, Duvinage, Mancas, Gosselin, & Dutoit, 2013), or to fully replace it by a hardware such as an eye tracker (O'Sullivan, Howlett, McDonnell, Morvan, & O'Conor, 2004).

The studies in perception has led to benefit computer graphics area by using perceptual quality metrics for improving the performance in areas such as Lighting, Rendering and Animation (Myszkowski, 2002) by reducing computational costs and use of perceptual metrics such as Animation quality metric (AQM) and Visible Difference Predictor (VDP). The many advantages of using perception to improve Interactive Graphics, Image Fidelity, Animation, Virtual Environment, visualization and Non-Photo realistic Rendering are highlighted by OSullivan et al. (2004). The perceptual use of impostor representation for simplification of human and building models in large scale crowd scenes and city scenes are evaluated and explored in Hamill, McDonnell, Dobbyn, and O'Sullivan (2005). The authors Gan, Cai, Liu, and Wang in the year 2015, similarly used perception based texture retrieval that are perceptually meaningful and non-uniform for non-expert users. Perception driven texture generation approach for generating textures with desired perceptual properties had been researched and proven to be successful (Gan et al., 2017). On similar lines, Perception-driven rendering methods has been studied to accelerate computational time for photo-realistic rendering to deliver 'full experience' to the end user without any compromise in visual quality by measuring perception metrics (Weier et al., 2017a).

Photorealistic rendering traditionally exploits perception limitations to accelerate costly light transport computations (Weier et al., 2017b) and in 3D graphics, HVS models allow removing nonperceptible components (Reddy, 1997, 2001) and/or predicting popping artifacts (Schwarz & Stamminger, 2009).

Perceptual models have been further applied to improving virtual simulations (Ondřej, Ennis, Merriman, & O'sullivan, 2016), character animations (O'Sullivan et al., 2004; Reitsma & Pollard, 2003), human body modeling Shi, Ondřej, Wang, and O'Sullivan (2017), fluid simulations (Bojrab, Abdul-Massih, & Benes, 2013; Um, Hu, & Thuerey, 2017), and crowd simulations (Wang, Ondřej, & O'Sullivan, 2017; Wang, Ondřej, & O'Sullivan, 2016). High dynamic range imaging and tone mapping benefits from models of human light adaptation (Ferwerda, Pattanaik, Shirley, & Greenberg, 1996; Mantiuk, Myszkowski, & Seidel, 2006), color to grey conversions simulate human color sensitivity (Neumann, Čadík, & Nemcsics, 2007; Smith, Landes, Thollot, & Myszkowski, 2008). Interestingness (Gygli, Grabner, Riemenschneider, Nater, & Gool, 2013) and aesthetic properties of photographs (Aydın, Smolic, & Gross, 2015), paintings and fractals (Spehar, Clifford, Newell, & Taylor, 2003; Taylor, Spehar, van Donkelaar, & Hgerhll, 2011) have also been approximated by computational models of HVS.

Close to our work is research on procedural textures (Liu, Dong, Cai, Qi, & Chantler, 2015b) that aims to define perceptual scales which can be used to steering texture model. The perceived quality of a geometry replaced with texture has also been studied (Rushmeier, Rogowitz, & Piatko, 2000).

Image quality metrics (IQM) utilize HVS models to predict perceptual image quality. Full-reference IQMs compute perceptual differences between the reference and distorted images (Mantiuk, Kim, Rempel, & Heidrich, 2011; Wang, Bovik, Sheikh, & Simoncelli, 2004; Wolski et al., 2018), while no-reference metrics (Herzog et al., 2012; Ye, Kumar, & Doermann, 2014) predict the quality in a reference-less setup. Video quality metrics (Aydın, Čadík, Myszkowski, & Seidel, 2010; Winkler & Mohandas, 2008) simulate temporal HVS properties to faithfully comparing video sequences. Recent research works study perceptual quality of 3D models (Lavoué, Larabi, & Vása, 2016) and models (Guo, Vidal, Baskurt, & Lavoué, 2015; Nader, Wang, Htroy-Wheeler, & Dupont, 2016) including textured models (Guo et al., 2016). Visual saliency predictors for 3D models have been also proposed (Wu, Shen, Zhu, & Liu, 2013).

Unfortunately, no existing metric is applicable to comparison of synthetic and real terrain images or models, because the compared contents differs significantly.

2.5 Perceptual Issues related to Procedural Modeling Area

The current procedural modeling methods try to address the need for modeling individual characteristics of model types and improving their quality ((Smelik et al., 2009) & (Smelik et al., 2014)). Liu, Dong, Qi, and Chantler (2013) was the first to address the need for identifying perceptual features of procedural textures. Liu et al. (2015a) also identified that the textures which are based on mathematical processes are not perceptually meaningful as they tend to be ”repetitive,” ”directional,” ”structured,” (Liu et al., 2015a). Zheng, Zhong, Liu, Cai, and Dong (2014) proposed a learning model and deep architecture for learning visual textures to understand their perception. Procedurally generated Cities and Buildings are also criticized because they realistic structure, lack of control, the generated methods are not of sufficient quality and the isolated feature set (Smelik et al., 2009), (Smelik et al., 2014) & (Musialski et al., 2013).

2.5.1 General factors that affect the perception of a model

a. 'Mesh Saliency' & 'Schelling Points'

'Mesh Saliency' is defined as the "measure of regional importance" of 3D models and (Lee et al., 2005) 'Schelling Points' is defined as the point of focus when perceiving an object in a graphically simulated scene (Chen et al., 2012). Both of these properties combinatorially, is very important in computer graphics applications as they aid in Object Recognition, Shape Matching, Shape-Based Retrieval, Metamorphosis, Cross-Parametrization, Texture Mapping, Deformation Transfer, Shape Approximation, View-Point Selection, Symmetry Detection and Part-Based Segmentation etc., of a 3D model (Chen et al., 2012). Therefore, from the perspective of my research topic, it is very important to determine the 'Mesh saliency' and 'Schelling points' of the specific category of the procedurally modeled model that will be studied in detail.

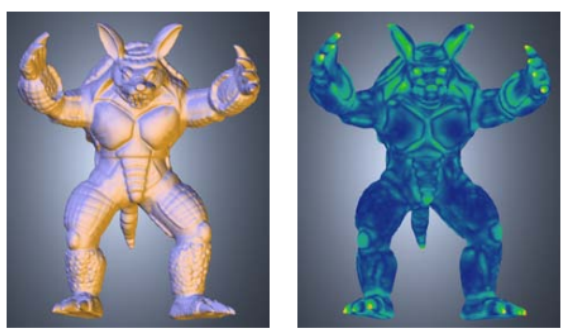


Figure 2.2. Armadillo model [Left] - Saliency Points [Right] (Lee et al., 2005).

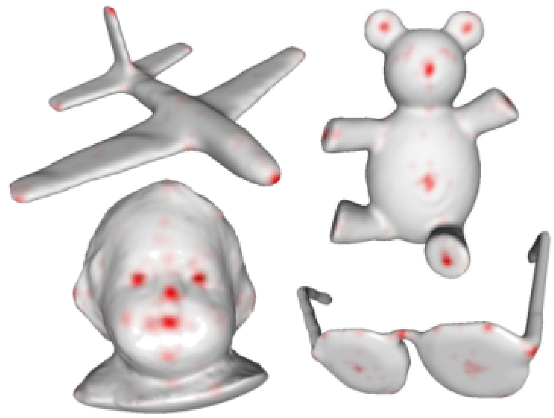


Figure 2.3. Schelling Points of 3D models (Chen et al., 2012).

b. Perspective (View point selection and Nice viewpoints)

”Perspective effects are pervasive in our visual worlds and provide a wealth of information about spatial layout” (Sedgwick, 1983) and (Thompson et al., 2011). The perspective of a scene that includes an object of interest will help the viewer in understanding the object size, distance, position of the object in relation to the figure-ground plane and many more visual details (Thompson et al., 2011).

But, not every perspective projection of a model in a scene can be categorized as a ’nice view point’. A nice view point of an object or a 3-D model as defined by (Toussaint, 2000) is a view that encompasses all the features of an object of interest with clear visibility (Kamada & Kawai, 1988). In addition to this, there are three different viewpoint quality measures: a. heuristic measure, b. view point entropy and c. Kullback-Leibler distance which is based on the projected and actual distribution of polygons (Neumann et al., 2005).

c. Model Texture

The texture of a model is the surface appearance of a model. The texture details of the model can give us many visual cues that correspond to the surface

Figure 2.4. "Representative views" of a 'Chair' model (Neumann et al., 2005).

orientation, material property and shape of the model itself ((Landy & Graham, 2004) and (Gibson, 1950)).

There are two different parent types of textures: a. Structural ("a specification of individual subpatterns making up a texture and an indication of how the subpatterns are replicated over the image region corresponding to the surface") (Thompson et al., 2011) and b. Statistical ("conditional distributions of nearby image locations or spatial frequency distribution")(Thompson et al., 2011).

d. Lighting (Illumination, Shading and Shadows)

The Illumination of an object indicated the source of incoming light of how the object is visible to the viewer. In Computer Graphics, there are three types of object illumination: 1. Direct Illumination, 2. Indirect Illumination and 3. Global Illumination based on the source of how the light arrives to surface of the object (Thompson et al., 2011). The continuous variation of the light formed on the surface of the object that is caused to due to the incident light reflection is called as 'Chiaroscuro' or 'Shading'. The light interacting with the surface of the object helps in revealing the chiaroscuro pattern that will help the viewer in evaluating the shape of the object through shading (Thompson et al., 2011). Objects tends to cast shadows on other objects on the scene when they are placed in a directional light field. The strength of the shadow will help in understanding the distance and shape of the object (Thompson et al., 2011).

Figure 2.5. Different types of Illumination as exhibited through a 'Cornell Box' model (Goral et al., 1984).

e. Material Properties

The type of material that an object is made up of helps in identifying several properties of the materials and how it interacts with the environment such as appearance, texture, light interaction and dynamics of the object (Thompson et al., 2011). In computer graphics, the material property of a model is usually defined through a shader system.

2.5.2 Specific factors that affect the perception of a certain model category:

In addition, to the above mentioned features in section 2.5.1, there are several unique identifiable features of a model that will affect the perception of its own model category. They are discussed in the next section for Terrains as follows.

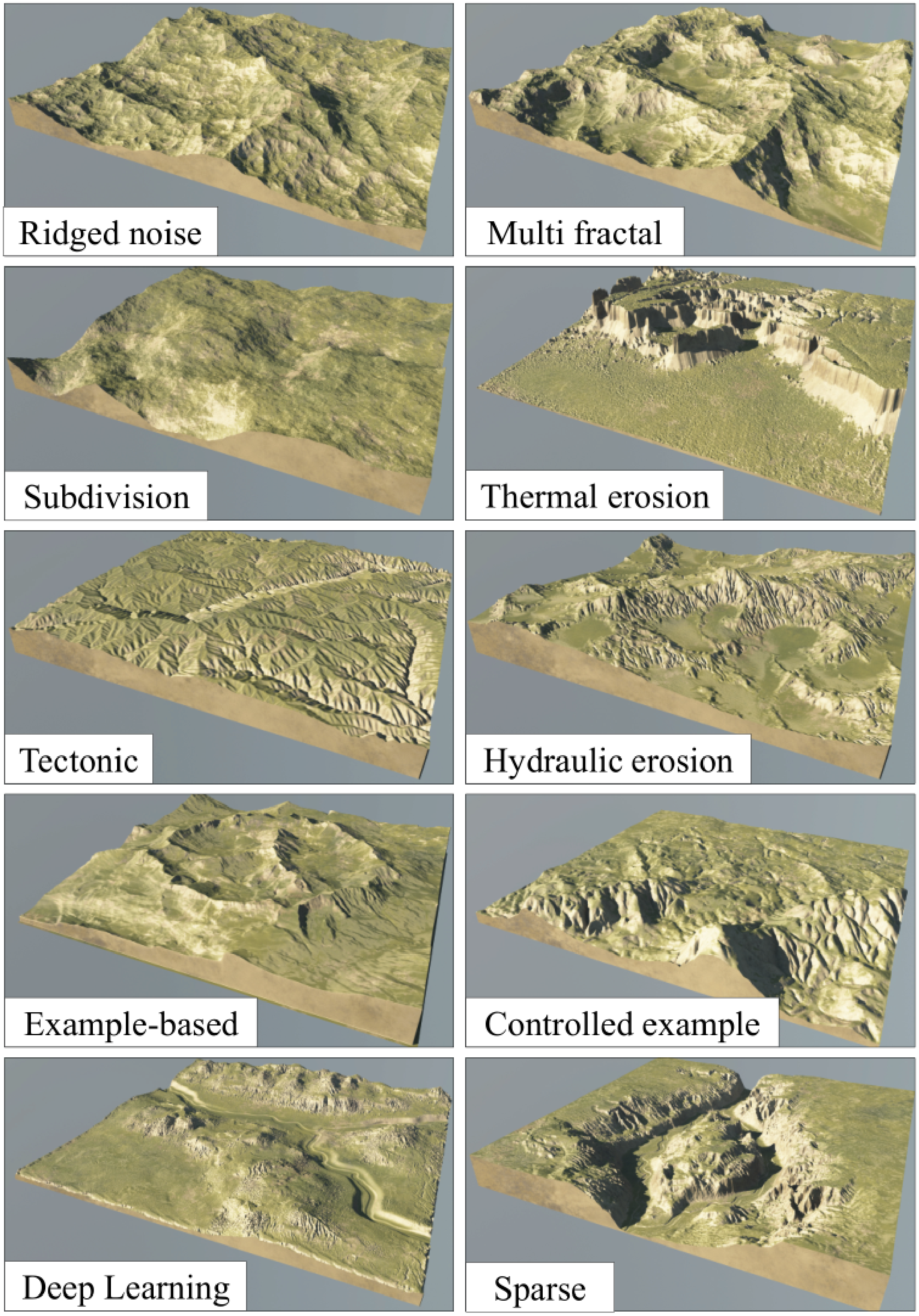


Figure 2.6. Examples of State of Art in Terrain Modeling in Computer Graphics as rendered by Galin et al. (2019)

2.6 Terrains

Terrains in computer graphics have been studied for several decades and a recent comprehensive review of generating methods can be found in Galin et al. (2019). Here we briefly list the three major categories of terrain generation techniques: procedural approaches, erosion simulation and from example synthesis.

Procedural Terrains in computer graphics are generated through the use of a 2-Dimensional height field in which the value of vertices represents the elevation at that location (Smelik et al., 2009) and (Smelik et al., 2014). The heightfield of a Terrain are generated through the use of stochastic methods (such as mid-point displacement, noise generation, fractal techniques), example based techniques (image or through the input of the user by sculpting) or simulation based approaches (Smelik et al., 2014).

Historically, the first methods to synthesize terrains relied on procedural and fractal approaches. It consists in finding a way to generate a fractal surface that exhibits self-similarity either by using subdivisions Fournier, Fussell, and Carpenter (1982); Miller (1986), faulting Mandelbrot (1988), or by summing noises Musgrave, Kolb, and Mace (1989). Approaches that control Kelley, Malin, and Nielson (1988) or more specific curve-based constructions Gain, Marais, and Strasser (2009); Hnaidi, Guérin, Akkouche, Peytavie, and Galin (2010) have been introduced. The overall realism of the generated landscape depends on the fine tuning of control parameters and requires a deep knowledge and understanding of the underlying generation process which restrict those methods to skilled technical artists. In contrast, erosion simulations aim at generating realistic terrain features by approximating the natural phenomena, such as hydraulic Benes, Těšínský, Hornyš, and Bhatia (2006); Krištof, Benes, Křivánek, and Šťava (2009); St'ava, Benes, Brisbin, and Křivánek (2008) or thermal erosion Benes and Forsbach (2002); Musgrave et al. (1989) processes at different scales.

In general, these methods are computationally intensive, and only capture a limited set of small scale structures features Cordonnier, Cani, Benes, Braun, and Galin (2018), such as ravines or downstream sediment accretion regions. When combined at a larger scale with uplift Cordonnier et al. (2016), erosion simulations generate realistic mountain ranges with dendritic ridge networks and their dual drainage network forming rivers.

Another option to obtain realism by synthesizing new terrains by-example, for example by stitching together terrain patches from existing data-sets. By using techniques from texture synthesis Gain, Merry, and Marais (2015); Zhou, Sun, Turk, and Rehg (2007) or sparse modeling Guérin, Digne, Galin, and Peytavie (2016), authors can generate large terrain with realistic small-scale features. The large scale plausibility remains an open challenge as existing methods, even deep learning Guérin et al. (2017) oriented approaches, rely on user-sketching and authoring.

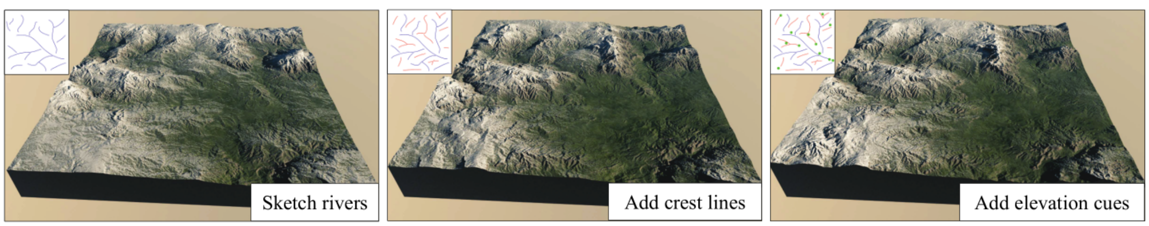


Figure 2.7. Interactive Example based Authoring of Terrains Guérin et al. (2017)

Despite recent advances in simulation, the user-control remains an open problem and terrain generation methods only generate a limited set of landforms. Moreover, validation of the generated structures remains an outstanding problem and has been addressed only partially.

2.6.1 Perception of terrains:

Synthetic terrains have not been studied in perception experiments and we are not aware of any computational perception quality metric that could be applied.

Furthermore, a data-set of synthetic and real terrain images comprising human judgments which could be used for an evaluation and comparison of terrain generating methods or for training of data-driven techniques is missing as well.

Nevertheless, a few research works on classification and perception of real-world terrains have been presented in the fields of environmental psychology and geomorphology. Dragut and Blaschke [2006] proposed a system for landforms classification on the basis of profile curvature. Several data layers are extracted from the digital terrain model to feed an image segmentation which classifies the terrain into classes like toe slopes, peaks, shoulders, etc. Fractal characteristics of terrains were studied in Hagerhall, Purcell, and Taylor (2004) and they conclude that there is a relationship between preference and the fractal dimension, meaning that fractal dimension may be part of the basis for preference. Finally, scenic beauty and aesthetics have been addressed by Daniel (2001); Palmer (2003); Tremblet (2016); Tveit, Sang, and Hagerhall (2012). These works lay the foundation of landscape perception, but they cannot be directly applied to quality assessment of synthetic terrains. Automated tools of measurement and analysis of terrains are sought Palmer (2003) to advance this area of research.

The inherent features of a Terrain model that the current procedural modeling methods address can be categorized as follows (Li et al., 2008), (Peytavie et al., 2009), (Smelik et al., 2009) and (Smelik et al., 2014):

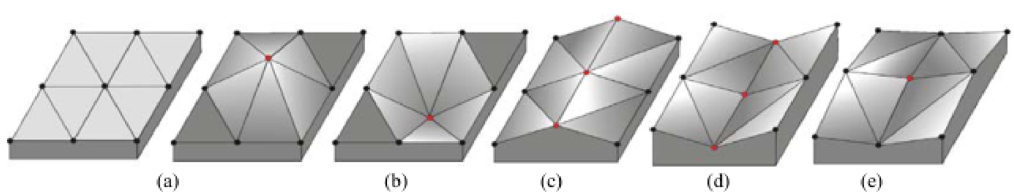


Figure 2.8. Categorization of Terrain Features (Li et al., 2008).

- Shape: Peak, Pit, Ridge, Channel, Pass.
- Layers: Loose Rocks, Overhangs, Arches and Caves.

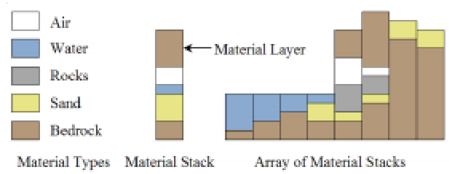


Figure 2.9. Material Layer and MaterialStacks (Peytavie et al., 2009).

- Feature: Lakes, Streams, Rivers and Oceans.
- Texture: Sand, Snow, Rocks and Grass.
- Modification of Appearance: Eroded Features (Aeolian, Hydraulic, Fluvial and Glacial).

The erosion (Wind, Water, Glacial and Sea) simulation of a Terrain and its features can be considered as its dynamic feature (Št'ava, Benes, Brisbin, & Křivánek, 2008), Vanek, Benes, Herout, and Stava (2011) and (Cordonnier et al., 2017).

2.7 Research Methods in Applied Perception and Computer Graphics

Qualitative research methods is one of the best answer for studying the phenomenon as their world view helps to "seek deep understandings of others unique socially constructed worlds and search for patterns in human behaviors" (Miller, 2005). The use of Qualitative methods in Computer Graphics are not necessarily new for studying perception. Hoffman and Nadelson (2010) used a mixed methods study for motivational engagement in video games. Dünser, Grasset, and Billinghurst (2008) demonstrates the variety of evaluation that have been used in a variety of Augmented Reality studies from the years 1993 to 2007. Interviewing participants and doing a thematic analysis is a common answer for detecting common perception attributes ((Dünser et al., 2008), (King et al., 2004) (Vaismoradi, Turunen, & Bondas, 2013)). Allen et. al in the year 2006 attempted to make sense of the barriers in IT for Women through the use of Interviews, performing Thematic Analysis on the collected data and demonstrating the results with the use of a causal maps. This seems very map-able to our problem of detecting perception issues in computer graphics scenes that uses procedurally generated models.

2.8 Machine Learning

The digital presentations of terrain are common as height maps in image format. We assume the real terrain height maps contain geomorphological elements that procedural ones do not have. One of our primary goals is to study the difference of the features between real and procedural terrains. To achieve this, we would like to apply the features from real terrains to procedural terrains, vice versa. The geomorphological features reflect in the images as structured feature pixels. Therefore, we can relate our work to Neural Generative Networks.

Recently, Machine learning has been picking up traction in computer graphics for solving a variety of problems without explicitly programming for a specific task through supervised, reinforced or unsupervised learning methods. In the year 2009, Judd et al predicted where the humans look in the screen through the use of a ML approach (Judd, Ehinger, Durand, & Torralba, 2009). A variety of applications and approaches for using Deep Learning in various scenarios has been thoroughly surveyed in 2015 by LeCun et al. (LeCun, Bengio, & Hinton, 2015a). Through the use of reinforcement learning, locomotion skills for an animation agent that adapts to various types of terrain has been found to be successful (Peng, Berseth, & Van de Panne, 2016). Data-driven shape analysis and processing has been studied thoroughly to provide suggestions on how to guide future studies that involves large data on how data is collected, feature extraction for learning, inference and reconstruction has been shown in detail by Xu et al. (Xu, Kim, Huang, Mitra, & Kalogerakis, 2016). We hypothesize that this technique will be very useful in our study for improving perceptual features in a 2D context.

Our work is based on recent advances in **deep neural generative networks**.Powered by Deep Learning LeCun, Bengio, and Hinton (2015b), a variety of applications and approaches were explored in various scenarios, especially with Convolutional Neural Network (CNN) on image understanding tasks He, Zhang, Ren, and Sun (2016); Krizhevsky, Sutskever, and Hinton (2012); Simonyan and Zisserman (2014). Besides the many breakthroughs with CNN such as object detection Redmon and Farhadi (2018), segmentation He, Gkioxari, Dollar, and Girshick (2017), and tracking Wang, Zhang, Bertinetto, Hu, and Torr (2019), the generative networks such as Neural Style Transfer (NST) Gatys, Ecker, and Bethge (2015) and Generative Adversarial Network (GAN) Goodfellow et al. (2014) belong to the most promising approaches that should find applications in computer graphics. The generative networks make pixel-wise alteration or synthesis for getting desired visual features in an image. Generative neural networks have been applied to terrain generation only in limited way Guérin et al. (2017). In this paper, we use generative networks to transfer important features from terrains perceived as realistic to non-realistic ones.

Recent NST work has been reviewed and summarized by Jing et al. (2017). The initial research of NST started with Image-Optimisation-Based (IOB) networks using parametric Gatys et al. (2015); Luan, Paris, Shechtman, and Bala (2017) and non-parametric methods Li and Wand (2016). Such methods can transfer artistic style or photorealistic features. However, the iterative image optimization procedure usually leads to low computational efficiency. The more recent NST direction shifts to Model-Optimisation-Based (MOB) methods, from per-style-per-model Johnson, Alahi, and Fei-Fei (2016); Ulyanov, Lebedev, Vedaldi, and Lempitsky (2016) to multiple/arbitrary-style-per-model architectures Chen and Schmidt (2016); Zhang and Dana (2018). Comparing to the online IOB optimizations, offline MOB models provide more efficient and flexible solutions. The per-style-per-model methods produce more impressive results close to Gatys et al. (2015) than the arbitrary-style-per-model methods while requiring more training efforts. The multiple-style-per-model methods are more suitable for the terrain feature transfer considering terrains with different erosion types as styles.

Generative Adversarial Neural Networks (GAN): Goodfellow et al. (2014) provides a neural network structure that utilizes a *generator* and a *discriminator* to adversarial synthesis of remarkable images. There is a large number of variants and enhancements of GANs that have been proposed and applied to different scenarios - from sketch to photo translation Isola, Zhu, Zhou, and Efros (2016) to face generator Karras, Aila, Laine, and Lehtinen (2017). A close work on terrain authoring Guérin et al. (2017) used conditional GAN Isola et al. (2016) to translate sketches into terrain representations. Several works Kim, Cha, Kim, Lee, and Kim (2017); Zhu, Zhang, et al. (2017) extended the conditional GAN Isola et al. (2016). The work of Cherian and Sullivan (2019); Zhu, Park, et al. (2017) enable domain translation without paired images in training.

The deep learning area is very quickly evolving but has been applied to the simulation of natural phenomena only in limited way. The goal of this work is to apply deep learning to transfer visually important features among terrains.

2.9 Geomorphology Based Evaluation Approaches

Huggett (2016) provides us with common landforms in the globe that are caused by many different types of erosions with corresponding patterns such as: Aeolian, Glacial, Fluvial, Slope, Coastal, Weathering etc.,. These landforms have similar landform features across them. Bullard and Livingstone (2002) suggests that there are interactions between these systems and they could not be viewed as mutually exclusive. Therefore, deriving a perceptual metric based on erosion patterns alone would not suffice and scale well. In the year 2013, Jasiewicz and Stepinski (2013) came up with a novel mapping and classification system of

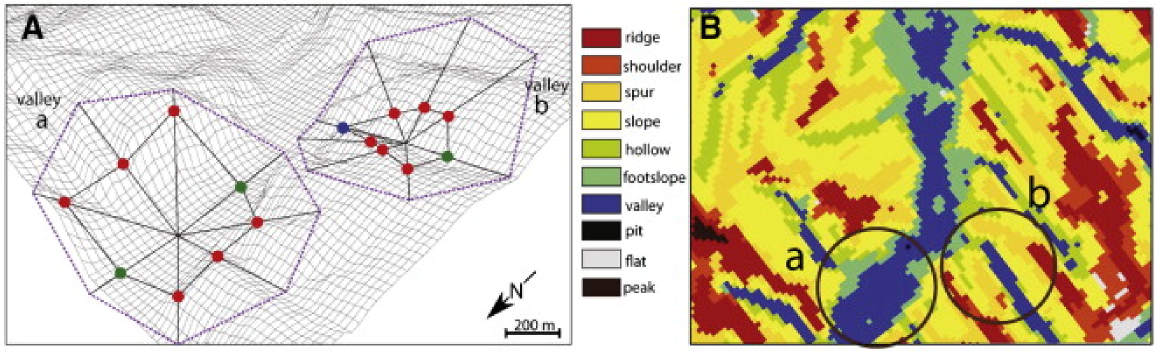


Figure 2.10. Machine Vision based feature classification provided by Geomorphons (image from Jasiewicz and Stepinski (2013)).

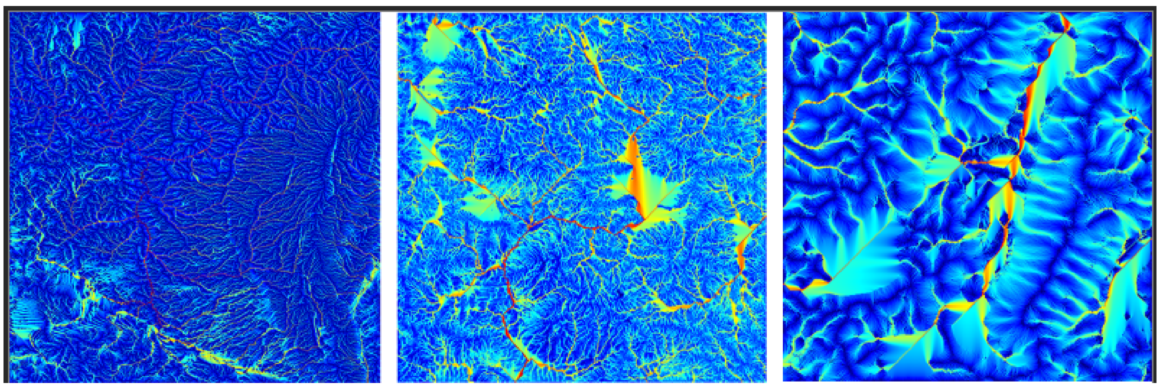


Figure 2.11. Hydrology based evaluation of Real and Synthetic Terrains.

landform elements for Digital Elevation Maps based on Computer Vision based pattern recognition approach. These landform will provide us with a basis for deriving a perceptual metric for terrain models in computer graphics. The geomorphons is discussed in detail in Section 4 of this research work. Although, We could also visually evaluate a terrain based on the flow path enforcement and filling but it only limits the evaluation category to Terrains with dendritic patterns Lindsay (2016). An example of such hydrology based evaluation for Real and Synthetic Terrains are given in 2.11.

2.10 Summary

The current procedural modeling methods though they address the needs for modeling individual characteristics of the aforementioned model types yet there is a research gap in this area for individual feature examination for their 'goodness' and perception with the exception of textures which is studied by Liu et al. (2015a). The procedural models such as procedural textures which are based on mathematical processes are not perceptually meaningful as they tend to be "repetitive," "directional," "structured," (Liu et al., 2015a). Additionally, there has been criticism of procedural city models such as "the cities they generate often lack a realistic structure" (Smelik et al., 2009) and "quality of the obtained results is not sufficient for most applications" (Musialski et al., 2013) for urban building models. Therefore, there is a need for detailed perception analysis of procedurally generated models that needs to be studied.

Unfortunately, the perception of procedural models (with the exception of Procedural Textures) in both static and dynamic scene context has not been studied in detail previously. This presents an opportunity for us to experiment and research for filling the research gap in the area that concerns Procedural Modeling and Perception analysis. The overarching idea will not only benefit in understanding the metrics and factors that lie behind perception of procedural models but also helping in improving the quality of generated models in procedural modeling by focusing on those specific metrics.

CHAPTER 3. METHODOLOGY

The main goal of undertaking this large scale perceptual experiment is to establish a set of visual perceptual attributes and in turn perceptual metrics for procedural terrain models. The perceptual attributes can be used as parameters to guide the neural network training for both identification and transfering features from real terrain models and procedural models. This chapter provides the methodology to be used in the research study for data collection, data analysis, feature extraction, learning and inference for our perceptual realistic reconstruction of the digital elevation model.

3.1 Method Overview

The key question we are trying to answer in this body of research is the visual plausibility of terrains and the visual perception and evaluation of synthetic terrains generated by terrain modeling methods in computer graphics. We focus on the terrain geometry only and we do not consider any additional features such as snow, vegetation, of water bodies. Our work builds on the recent advances in geomorphology, in particular we use the concept of *geomorphons* that are features extracted from Digital Elevation Models (DEMs) that quantitatively measure presence of various shapes in terrain (Section 4.1).

3.1.1 Research Method

We performed a large scale user study in two-phases (Section 4.2) in which we quantify the perception of real and synthetic data-set and the transferred features. Figure 3.2 shows the overview of our testing. Our goal is to quantify the levels of perception on a given model. On a higher level, a web based survey with two rendered images of models (Procedural Models vs Models from other means) side by side and asking the samples to choose which image is realistic were help us in identifying their preferences seemed like a direct approach. A rendered image is better than showing the research participants, an interactive 3D Model because in real life most of the participants are going to have experience with a procedural model in a certain context (2D) that is not tangible (3D). A cognitive web-based survey was conducted to accomplish the study objectives. The participants in the survey were asked to choose between two set of images based on their realism. The investigators imposed minute variations in the procedurally generated terrains when compared with their natural counterparts. Finally, based on the survey results, the list of parameters that the audience has chosen were be composed and used for completing the objective of the study. The main goal of doing this study is validating the outputs of the proposed procedural natural object generation model by showing the variety of generated models to the audience.

During the *initial data generation*, we acquired data of real terrains from Shuttle Radar Topography Mission and we carefully selected several classes featuring prevalent geological patterns (see Table 4.1): Aeolian, Coastal, Fluvial, Glacial, and Slope (4.1). Then we generated synthetic data-sets by using terrain generation algorithms used in computer graphics: coastal, thermal and fluvial erosion, fractional Brownian motion, noise and ridged-noise terrain models. Geomorphons were generated for each image.

3.1.2 Measure for Success

The measure of success for the research experiments is defined by the successful identification of attributes that are influential in 'realistic' visual perception, classification of procedural Terrains models.

Description of Experiment-α (E-α) was our initial experiment with 10 Real Terrains (AEOLIAN Patterns x4, FLUVIAL Patterns x3 and GLACIAL Patterns x3) and 10 guided Synthetic Terrains with state of the art erosion algorithms. This was a pair-wise perceptual comparison with two-alternative forced choice design – 2AFC in Amazon Mechanical Turk platform. This test served more as a proof of concept and validated our concerns that Real Terrains perceptually rank higher than the Synthetic Terrains. we asked our test participants the question: *"Which image looks better: The HIT consists of two images which are marked as "Left" and "Right". Please choose between "Left" or "Right" images depending on which image you think looks the best?"*. The real terrains received 11,311 and synthetic terrains received 5,558 votes out of 16,869 total votes by 348 unique participants in the survey.

We performed statistical tests on our normalized perceptual scores to determine if there are any differences in perception of our terrain data groups in E-α: REAL$_{\alpha}$ vs SYNTHETICS$_{\mathrm{E}\ \alpha}$. We state the null hypothesis, H_0 for our statistical test in E-α as follows: *"There are no significant differences in the visual perception scores between our terrain data groups."*. We used the significance level of $\alpha = 0.01$, and get the statistics for, REAL$_{\alpha}$ versus SYNTHETICS$_{\alpha}$ ($p - value < 0.01$, $DF = 3373, t = 2.32$). It is highly evident from the test that Real Terrains are perceptually ranked than the Synthetic Terrains proving our notion. Therefore, we can reject the null hypothesis and proceed with further tests.

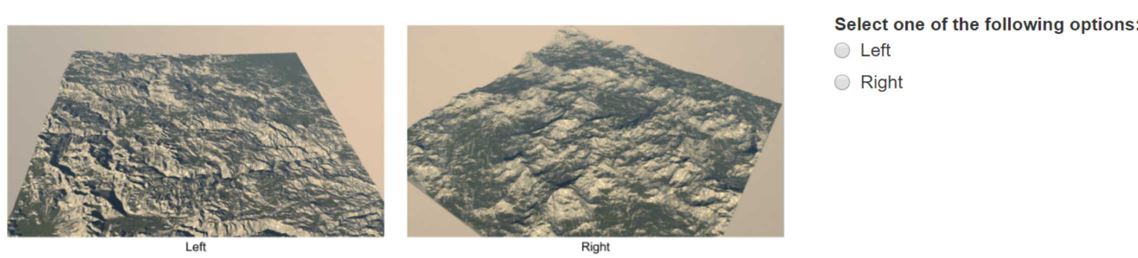


Figure 3.1. A Sample HIT of Experiment α in the Amazon Mechanical Turk test Platform

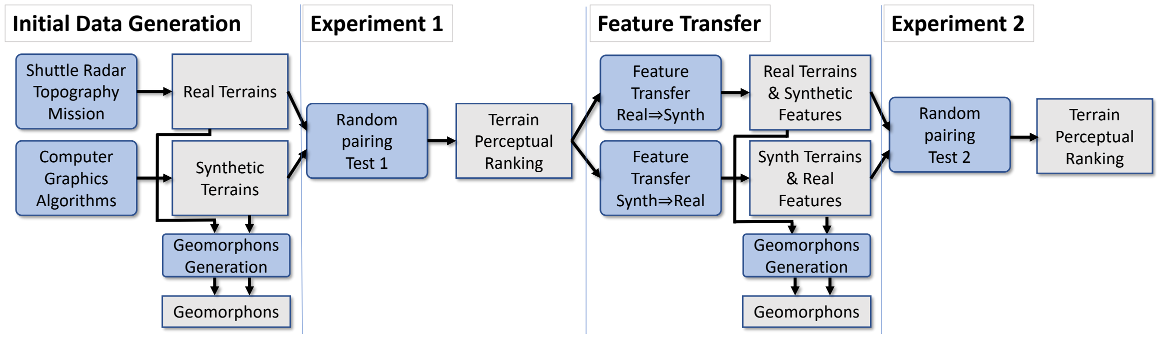
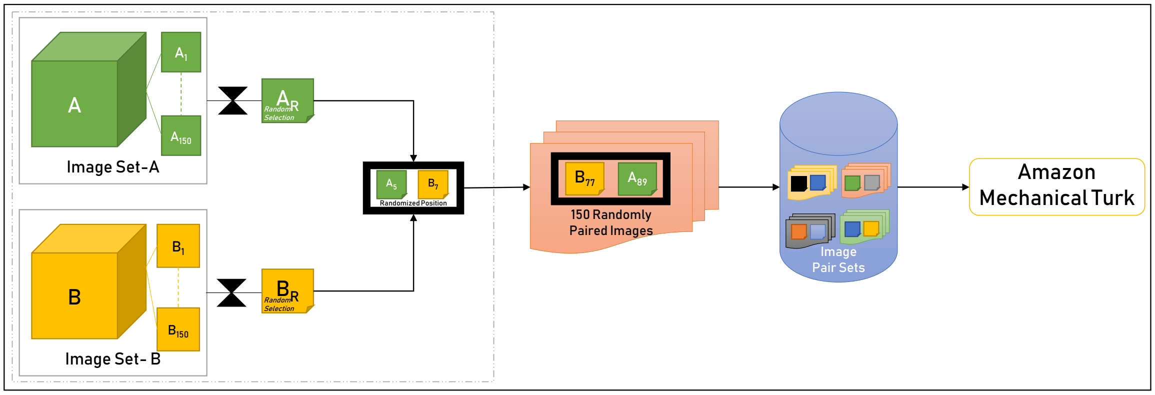


Figure 3.2. Overview of our experiments (boxes with rounded corners describe processes and squared boxes describe data). The *initial data for Experiment 1* were acquired from two sources: real and synthetic data, they were rendered and we also generated geomorphons for each image that quantitatively describe the landform features it contains. During the *Experiment 1* we acquired perceptual ranking of each image. During the *feature transfer* we transferred features from highly ranked images (Real⇒Synth) and vice versa (Synth⇒Real) resulting in two new datasets. During the *Experiment 2* we perceptually evaluated the initial data and the newly generated ones, confirming that the transferred features have visual importance on the perceived visual quality.

Description of Experiment-1: (E1) The first user study is a pair-wise perceptual comparison (two-alternative forced choice design – 2AFC) by using Mechanical Turk. The experiment provided initial terrain ranking for each image and for each image category within each group. The random pairing is based on the process shown in 3.3. In particular, we have shown pairs of images and we asked the viewers the question: *"Which terrain looks more realistic (left or right)?"*. Each image received multiple rankings and the number of votes decided its positioning in the overall test.

Feature Transfer: After the first user study we used the CycleGAN (Zhu, Park, et al., 2017) to transfer features from the images that were ranked high to those ranked low and conversely (Section 4.2.2). The motivation for this step is the underlying assumption that certain features have important effect on the visual perception of terrains. This step generated a new data-set that we call S2R (synthetic to real) and R2S (real to synthetic). S2R indicates that procedural



	R	S	R2S	S2R
R	R-R	R-S	R2S-R	S2R-R
S		S-S	R2S-S	S2R-S
R2S			R2S-R2S	S2R-R2S
S2R				S2R-S2R

Figure 3.3. (Top) An overview pipeline of our Image random pairing process across our experiments: Experiment 1 & Experiment 2. The subscript-'R' indicates a randomly chosen image from a set e.g., A_R indicates a randomly chosen image from Image set A. (Bottom) This figure shows the combinationsof sets used for generating our pairwise comparisons used in the Experiments

features were transferred to the real terrains and R2S is the opposite process. Geomorphons were also generated for the new data-sets.

Description of Experiment-2: (E2) The second user study is the same as E1, but also included the newly generated sets (Section 4.2.3). The random pairing is based on the process shown in 3.3. The underlying assumption was that the features from the highly ranked terrains were be transferred to the low ranked terrains and the new terrains were improve their ranking. The same expectation was hold for the terrains ranked in the opposite way. Moreover, for each terrain we also generated

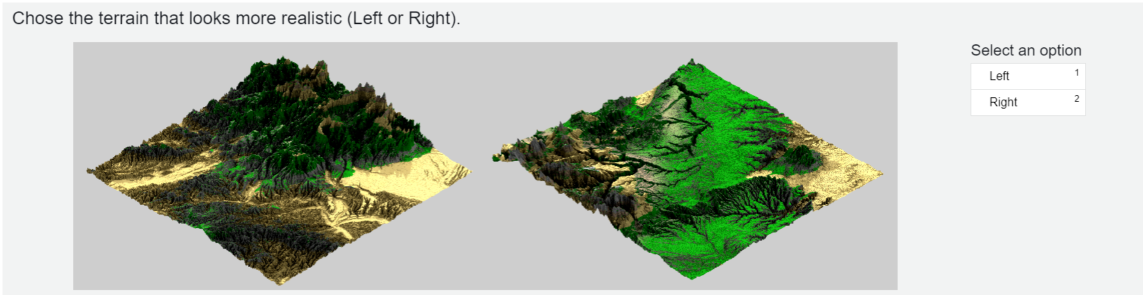


Figure 3.4. A Sample HIT of Experiment 1 in the Amazon Mechanical Turk test Platform.

the corresponding geomorphons and we kept a careful track of which features were transferred.

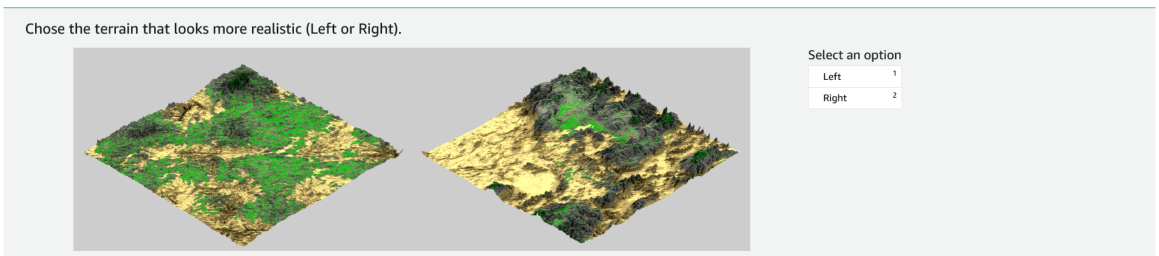


Figure 3.5. A Sample HIT of Experiment 2 in the Amazon Mechanical Turk test Platform.

3.2 Summary

This chapter provided the framework and methodology to be used in the research study. The next chapter will discuss our implementation details, neural network setup and style transfer process in detail.

CHAPTER 4. EXPERIMENTS

4.1 Terrain Data and Geomorphons

The objective of these studies was to compare the terrains with the most prevalent geomorphological processes with visually distinguishable features and the common terrain synthesis methods in computer graphics. The DEMs used in this study come from Shuttle Radar Topography Mission (SRTM) research data-set Farr and Kobrick (2000). We used the three arc-second capture resolution (90m pixel resolution along the equator) tiles from the data set as some of the tiles from the one arc-second (30m pixel resolution along the equator) data that covers the whole globe is not made available to public yet. The resolution roughly translates to 1° Longitude $\times 1^\circ$ Latitude or 100×100 km resolution approximately depending on the DEM's location on Earth. All the DEMs we used maintained a resolution of 512×512 that gives sizes of the land features around 200 meters per pixel. We determined 512×512 image resolution for height maps to be a nice balance in terms of visual features and image size for neural network training based on Peak signal-to-noise ratio [PSNR] estimation in comparison with other resolutions such as 128×128, 256×256, 768×768, 1024×1024 and 1201×1201.

4.1.1 Real terrains

Without loss of generality we used terrains that include patterns that commonly results from aeolian, glacial, coastal, fluvial, and slope processes Huggett (2016) along with the retrievability of suggested patterns from the SRTM data-set Farr and Kobrick (2000). It is important to note that the geoforming processes are not well-understood and most of the terrains are affected by several of them either at the same time period or in an indeterminable unknown sequence. So instead of discussing processes, we consider terrains that include the specified geomorphological patterns more or less. The two top rows of Figure 4.1 show examples of several renderings of real terrains and the supplementary materials include all data.

4.1.2 Synthetic Terrains

Again, Without loss of generality we used terrains generated by noise Perlin (1985), ridged-noise, fractional Brownian motion (fBm) surfaces Fournier et al. (1982), thermal erosion Musgrave et al. (1989), fluvial erosion Anh, Sourin, and Aswani (2007); Krištof et al. (2009); Neidhold, Wacker, and Deussen (2005); St'ava et al. (2008), and coastal erosion approximated by hydraulic erosion applied only to coastal areas (see Galin et al. (2019) for an overview). Eroded terrains were generated from noise-based terrains (Figure 4.1 two bottom rows).

While procedural generation of terrains is simple so we could have generated an arbitrary number of DEMs, it is rather difficult to find good samples for all the above-mentioned examples of real patterns. Table 4.1 shows how many terrain models we had for each category and also establishes nomenclature for each set. Each real image starts with letter R and synthetic with S, the second letter indicates subcategory. We refer to all images from real dataset, R and all synthetic as S. The size of each data-set was the same: $|S| = |R| = 150$.

4.1.3 Rendering

All terrains were rendered by using exactly the same settings to avoid any bias.

The camera position was set to display the terrain from about 45 degrees angle that is a common viewing distance from a top of a mountain or a low flying aircraft. This location shows enough details as opposed to top view and does not

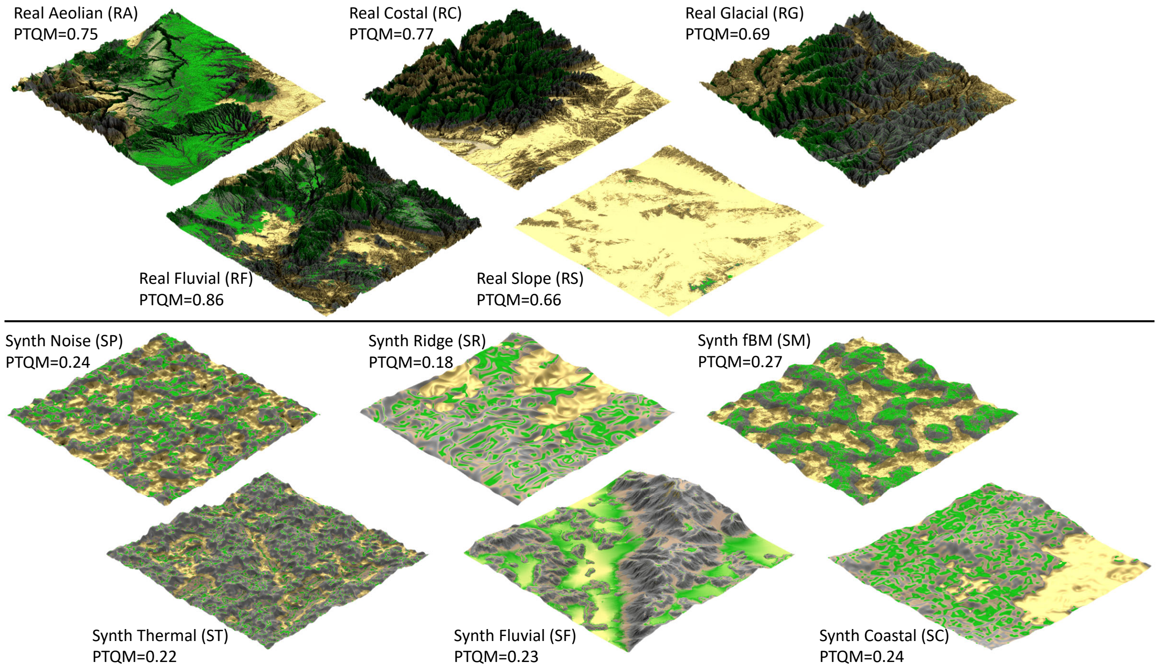


Figure 4.1. (Top) Examples of real terrains rendering used in our experiment and their PTQM: RA) aeolian patterns from Moab Arches National Park Utah USA, RC) coastal patterns from Gobi desert Mongolia, RG) glacial erosion patterns from Himachal Pradesh Western Himalaya India, RF) fluvial pattern from Chichiltepec Mexico Guerrero, and RS) slope pattern from Death Valley California USA. (Bottom) Examples of synthetic terrains SP) noise-based, SR) ridged-noise, SM) fractional Brownian motion surface, ST) thermal erosion, SF) fluvial erosion, and SC) coastal erosion (see supplementary material for high-resolution images).

cause self occlusions as opposed to side view. We also positioned the camera above one of its randomly chosen corners. We assumed viewers are familiar with this viewing angle.

We used sky sphere for illumination with gradient from 50% of gray near the horizon to full white in zenith. The rendering was performed by using global illumination with no additional lights, by using 500 reflections of light and 9× super-sampling for anti-aliasing. Each terrain was textured by the same color map that changed from low-level and flat areas with yellow color (sand), medium levels flat green (grass) to high and steep slopes gray (stone). We experimented with pure grayscale rendering by using only ambient occlusion, but the consistent color

Table 4.1.

Terrain type (real/synthetic/transferred features), categories, abbreviations, and the number of terrain samples in each category.

Type	**Category**	**Abbr.**	**Sampl.**
Real (R)	Aeolian	RA	55
	Coastal	RC	19
	Fluvial	RF	64
	Glacial	RG	07
	Slope	RS	05
Synthetic (S)	Coastal	SC	25
	fBm	SM	25
	Fluvial	SF	25
	Noise	SP	25
	Ridged-noise	SR	25
	Thermal	ST	25
Transferred	Synth features to real terrains	S2R	150
Features (2)	Real features to synth terrains	R2S	150

mapping provides better contrast of the final images. We intentionally used non-photo-realistic Gooch, Gooch, Shirley, and Cohen (1998) rendering so as to avoid any bias introduced by the simulation of vegetation and realistic rock, sand or grass rendering. Moreover, non-photo-realistic rendering enhances the shape and structure of the bare elevation of terrain which is the focus or this study.

4.1.4 Geomorphons

The fundamental theory behind our method is the recently introduced concept of *geomorphons* Jasiewicz and Stepinski (2013) that provide an exhaustive classification of terrain features from digital elevation models (DEMs) based on pattern recognition approach. Geomorphons build on the decomposition of DEM into local ternary patterns Liao (2010) that generate an oriented eight directional feature vector for each location of the DEM; one value for the Moore neighborhood. This gives rise to ten geomorphons (flat, peak, ridge, shoulder, spur, slope,

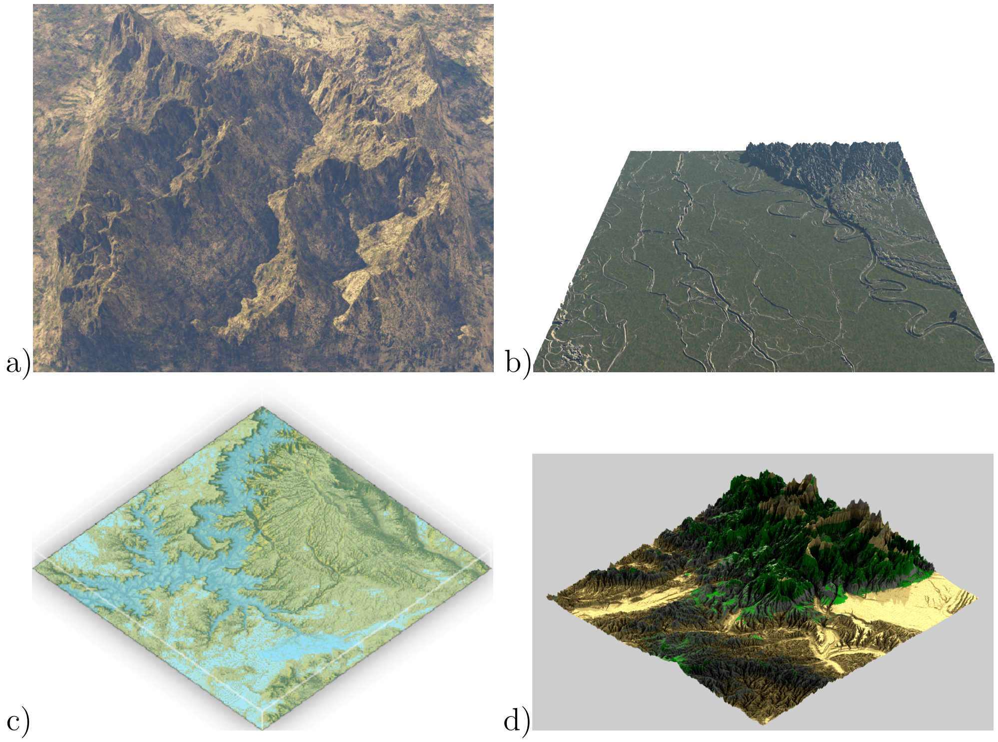


Figure 4.2. Examples from our different rendering experiments in tools such as: a) Terragen, b) Vue, c) RayShade module in R, d) POVRay.

depression (or pit), valley, footlsope, and hollow) as shown in Figure 4.3 from Jasiewicz and Stepinski (2013).

We used an open implementation of geomorphons in GRASS GIS tool Neteler and Mitasová (2013) that generates color-coded image corresponding to the input DEM as shown in example in Figure 4.4. The output of the algorithm is the percentage of coverage of each geomorphon in the input DEM.

We utilize geomorphons in the context of understanding the importance of individual geomorphological landform features and how they affect the perception of terrains. In the next text we show how they are present or missing in different terrains. The order of the geomorphons in the color coding in Figure 4.4 is arbitrary. In order to compare the wide variety of terrains used in this paper, we decided to sort the geomorphons according to their presence in the best visually

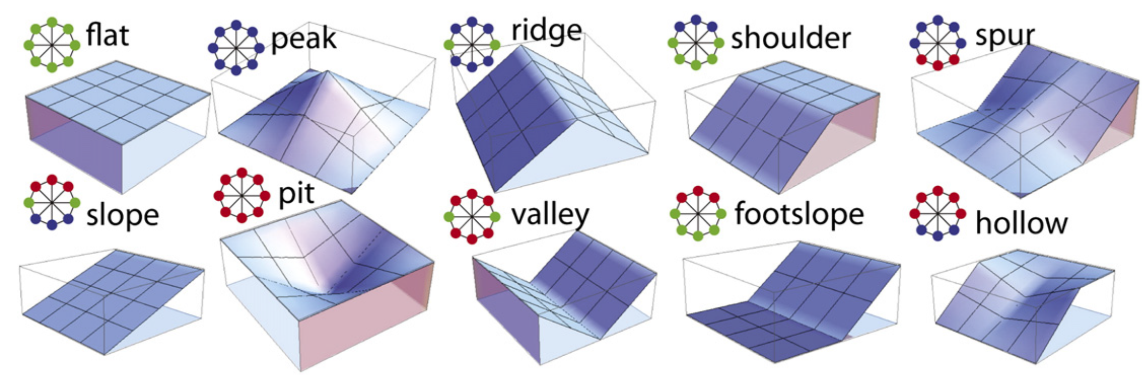


Figure 4.3. Ten most common land form patterns can be uniquely classified by geomorphons from a DEM. Blue disc identify lower, red higher, and green the same altitude (image from Jasiewicz and Stepinski (2013)).

perceived terrain category from our user study that are glacial patterns of real terrains (Section 4.2.1). Figure 5.11 shows the normalized frequency of geomorphons in all dataset used in this paper and we use the ascending order of geomorhons as: *Depression (or pit)* (the least present), *Summit, Flat, Valley, Ridge, Hollow, Spur, Shoulder, Slope, and Footslope* (the most frequently present). Please note that the values of geomorphons for all dataset from this paper will be made available as a supplementary material.

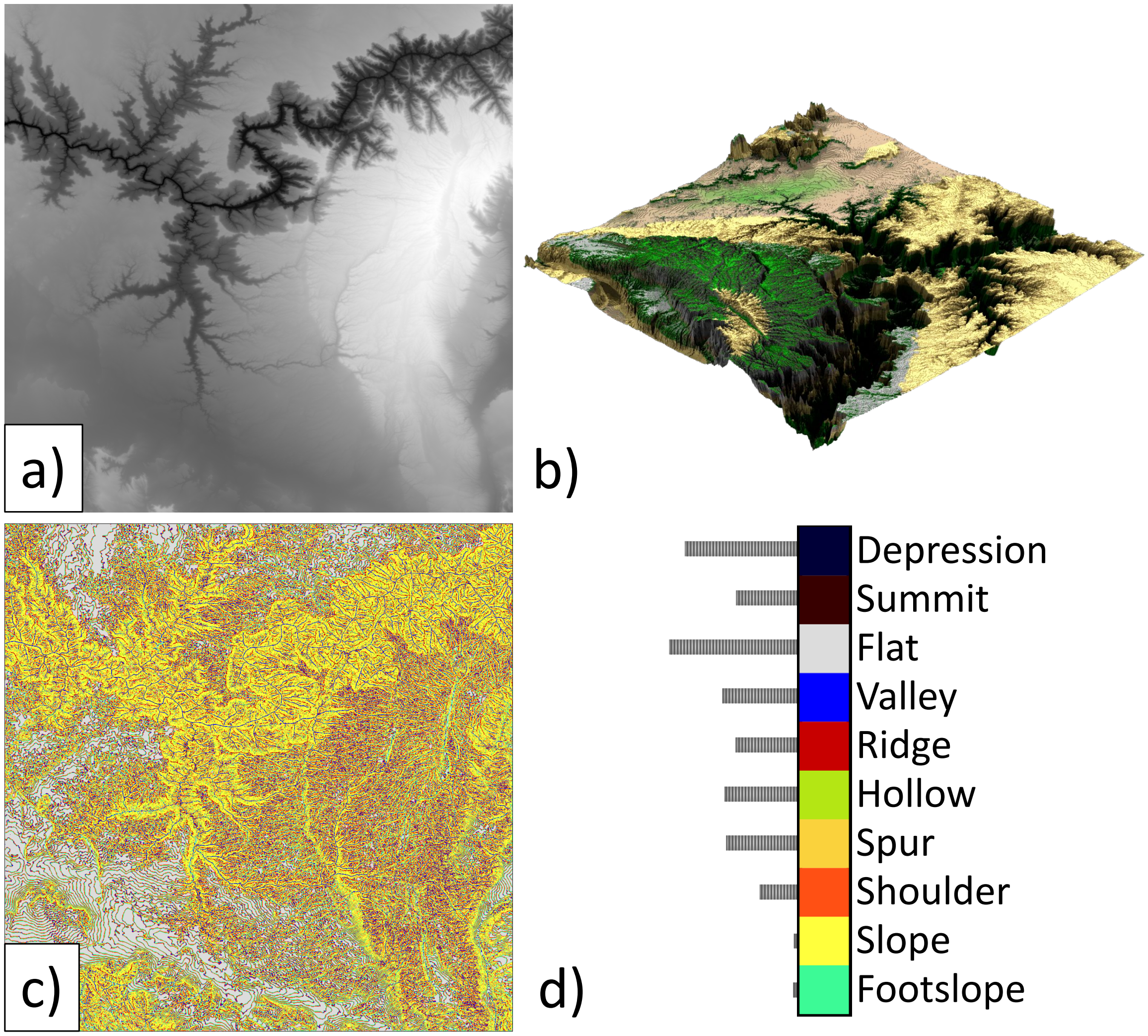


Figure 4.4. a) The input DEM b) its rendering and c) the geomorphons d) with the explanation of the color-coding.

4.2 Perceptual Experiments and Feature Transfer

The perceptual study was run on the Amazon Mechanical Turk platform and we asked the subjects the question: *"Which terrain looks more realistic (left or right)?"* by a showing a pair of terrain images without giving any other information about the terrain. Only qualified "Mechanical Turk Masters" were allowed to answer the survey and each repetitions of the image pair during the survey process were restricted to be answered only by unique participants every time. The survey is blinded such that the participants only see an image pair with responses restricted to 'Left' or 'Right' option. The experiment involved 70 participants with no particular constraints on their education or previous knowledge. All participants were older than 18 years. The source category information of each terrain that is being shown for collecting responses would not be displayed and randomized every time.

For each image pair we denote the category by dash, so R-S indicates pair of images where one is from the real and one from the synthetic sample. The actual position of each image (left or right) was randomized that makes this relation symmetrical: R-S is the same as S-R.

4.2.1 Experiment 1: Perceptual Evaluation of Real and Synthetic Terrain Models

We generated random image pairs for our 2AFC study by using the rendered images from Section 4.1. We randomly paired one real with one synthetic resulting in 150 image pairs. This pairing happened five times for each image from R resulting in |R-S| =750 images.

1. we randomly paired one synthetic with another synthetic terrain (another 150 images called S-S) and we repeated this step five times for each S resulting in |S-S| =750.

2. Also, one real image was paired with one real (150 image pairs called R-R) repeated five times resulting in |R-R|=750 image pairs. The total number of image pairs for E1 was |R-S|+|R-R|+|S-S| = 2,250 image pairs.

We made sure that the pairing did not miss any image, each image was repeated exactly five times, and pairing occurred always with a different image. The order of the images within each pair was also randomized so that the synthetic image could be on the left hand or right hand side of the pair with the same probability.

During the experiment each image pair was shown to five different participants who chose to participate in the survey resulting in a total 3750 image

Table 4.2.

The configuration for our Experiment 1 (E-1).

	TEST-1
Datasets Used	Real (R), Synthetic (S)
Count of Terrains	R = 150, S = 150
Across Comparisons	R-S (150)
Within Comparisons	R-R (150), S-S (150)
Unique Terrains	450
Image Pair Count after random matching	2250
Repetitions	5
Total Votes	11250

pair observations by a total of 70 subjects (out of which 33 respondents participated in responding for Across comparisons: R-S) with varying degree of participation (determined based on the unique count of anonymized 'workerID' provided by Amazon Mechanical Turk).

Each time an image was selected as more realistic, it received a point, and the total number of points determined the overall ranking of each image that was normalized as discussed in chapter 5. Moreover, we also calculated the normalized ranking of each category of real (RA, RC, RG, RF, and RS) and synthetics (SP, SR, SM, ST, SF, and SC) terrains.

4.2.2 Feature Transfer

E1 provided a ranking of each category of real and synthetic terrains. We observed that the real terrain contained features that the synthetic ones do not have, the valley topology in the terrains with fluvial erosion. Moreover, within the real terrain domain, different erosion causes different feature patterns, reflecting in a height map as patches of pixels with varying distributions of elevations.

We assumed that a deep neural network could learn the features that make real terrains visually plausible and that such features can be transferred onto the synthetic terrains to make them more visually realistic.

We initially experimented with neural style transfer (NST), which did not perform well, because the model lacks the capability of transferring consistent global topologies such as long ridges or valleys as discussed by Gatys et al. (2015). Because explicit pairing between the real and synthetic terrains is difficult, we consider using the unpaired image to image translation Zhu, Park, et al. (2017) to transfer features from the real domain to the synthetic domain, and vice versa. Moreover, we also hypothesize that the same transfer could be used to diminish features if the transfer occurs from synthetic to real terrains that would justify the importance of specific features.

The pipeline with major components of our network is shown in Figure 4.5. We use a pair of generators G_R and G_S with a pair of discriminators D_R and D_S. The generator G_R translates terrains from the synthetic domain S to the real domain R with real features. The discriminator D_R discriminates between terrains $\{r\}$ and $\{G_R(s)\}$, where $\{r\} \in R$ and $\{s\} \in S$. Moreover, G_S translates terrains within the real domain R to the synthetic domain S with synthetic features. Similarly, D_S discriminates between terrains $\{s\}$ and $\{G_S(r)\}$, where $\{r\} \in R$ and $\{s\} \in S$. Besides the adversarial loss, a cycle consistency loss is used to make $G_S\left(G_R(s)\right) \approx s$ and $G_R\left(G_S(r)\right) \approx r$. This process is indicated with the dashed arrows in Figure 4.5. The cycle consistency ensures the high-quality feature transfer.

We adopt a nine res-block generator and a 70×70 PatchGAN discriminator Isola et al. (2016). The checkerboard patterns (Figure 4.6) happen in the transferred terrains caused by fractionally-strided convolution and the artifacts decrease if the training epochs increase. We also applied resize-conv with Nearest Neighbor and Bilinear as suggested in Odena, Dumoulin, and Olah (2016) (Figure 4.6).

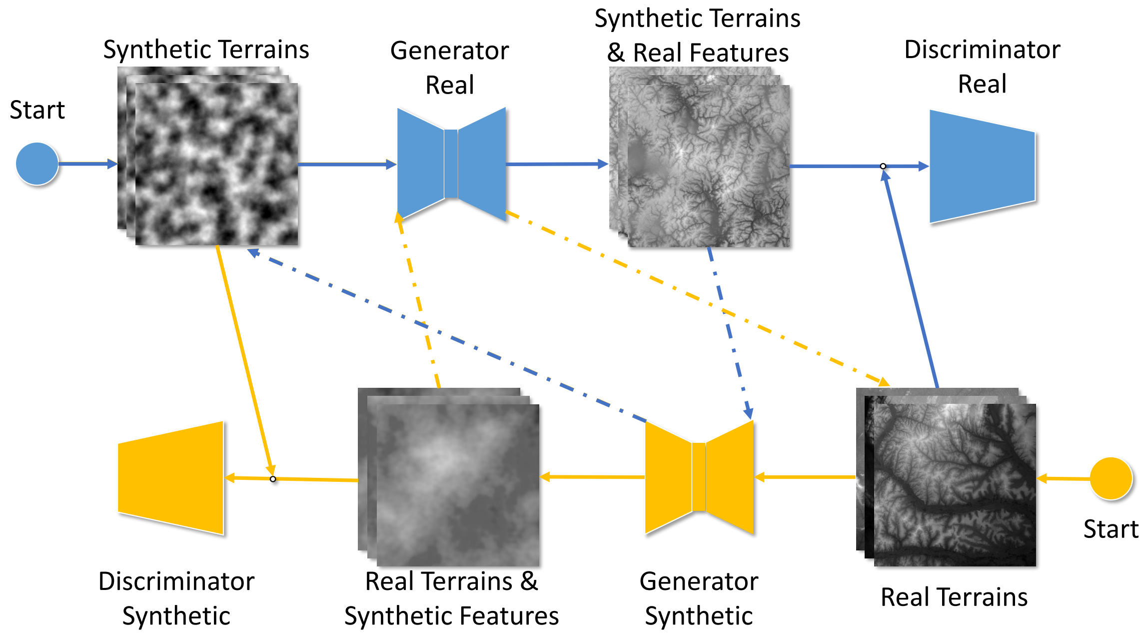


Figure 4.5. The pipeline with major components of the network for feature transfer. The blue arrows indicate the working flow of R2S; the orange arrows indicate the working flow of S2R. The dotted-and-dashed arrows indicate the cycle consistency process.

Our training set contains $9,800$ real terrain height maps selected from the SRTM DEMs excluding the terrains that have been used in our E1 and E2. We generated additional synthetic height maps for use in training based on aforementioned synthetic categorization and same size as the real terrain training data which is $9,800$ (see the data collection in Sections 4.1.1 and 4.1.2). Note that the sets from the first experiment were also included in the second one. In this way we have validated the first experiment, because the ranking of the results was consistent between E1 and E2 (Section 5).

We trained the model with 20 epochs, and have then generated 150 images of real terrains with synthetic features denoted by S2R; the notation denotes the *transfer* meaning "synthetic features were transferred to real terrains" and we also generated another 150 images of synthetic terrains with real features denoted by R2S. Figures 5.1 and 4.7 shows example result of the feature transfer in both directions (from real to synthetic and from synthetic to real) and Section 5 further discusses results.

Initially, we also experimented with another neural style transfer (NST), which did not perform as well as GAN. We based our NST on a multiple-style-per-model net Zhang and Dana (2018), considering a) a 2D style embedding (CoMatch) layer can enrich the representation of terrains with different erosion types (styles), b) a controllable brush-size supports varying density erosion features, and c) a good performance. Local fractal features in the real terrains can be well transferred to the procedural ones. However, the model lacks the capability of transferring consistent global topologies such as long ridges or valleys. As discussed by Gatys et al. (2015), large scale structures in the images usually response in deep layers. However, by adding hidden layers to the transformation network of NST, we did not observe significant improvement in transferring global geomorphological processes in our experiment. We may further improve the quality of the images by using unsupervised Almahairi, Rajeswar, Sordoni, Bachman, and Courville (2018) or supervised Zhu, Zhang, et al. (2017) approaches for different erosion types in the future work.

4.2.3 Experiment 2: Perceptual Evaluation of Real, Synthetic, and Terrain Models with Transferred Features

The objective of the second experiment (E2) was to evaluate how the terrains with transferred features score perceptually against real and synthetic terrains. We have reused the 750 R-S image pairs from E1 (Section 4.2.1) and added another 750 images for each missing combination. Table 4.4 shows the naming of the image pairs. The first column shows the reused pairs from E1 (R-S). The newly added pairs compare newly created transferred features from synthetic to real R2S combined with all options R2S-R, R2S-S, and S2R-R2S. Also, we added combinations for feature transfer from real to synthetic S2R S2R-R and S2R-S.

R2S-S2R is already included because it is symmetrical with S2R-R2S. Again, each shuffling was generated five times resulting in 750 images for each item of Table 4.4 resulting in total of 4,500 image pairs. We have repeated each test for five independent viewers and this resulted in the total of 22,500 views by 128 subjects (out of which 128 respondents participated in responding for Across comparisons: R-S, R-R2S, R-S2R, S-R2S, S-S2R, R2S-S2R). All participants were older than 18 years. Note that because the R-S set from the first experiment were also included in the second one, we have validated the first experiment, because the ranking of the results was consistent between E1 and E2 suggesting the data saturation point has been attained. (Section 5).

Table 4.3.

The configuration for our Experiment 2 (E-2).

	TEST-2
Datasets Used	Real (R), Synthetic (S), Real Features to Synthetic Terrains (R2P), Synthetic Features to Real Terrains (P2R).
Count of Terrains	R = 150, S = 150, R2P = 150, P2R = 150
Across Comparisons	R-S (150), R2P-P2R(150), R2P-R (150), P2R-R (150), R2P-S (150), P2R-S (150)
Within Comparisons	R-R (150), S-S (150), R2P-R2P (150), P2R-P2R (150)
Unique Terrains	1500
Image Pair Count after random matching	7500
Repetitions	5
Total Votes	37500

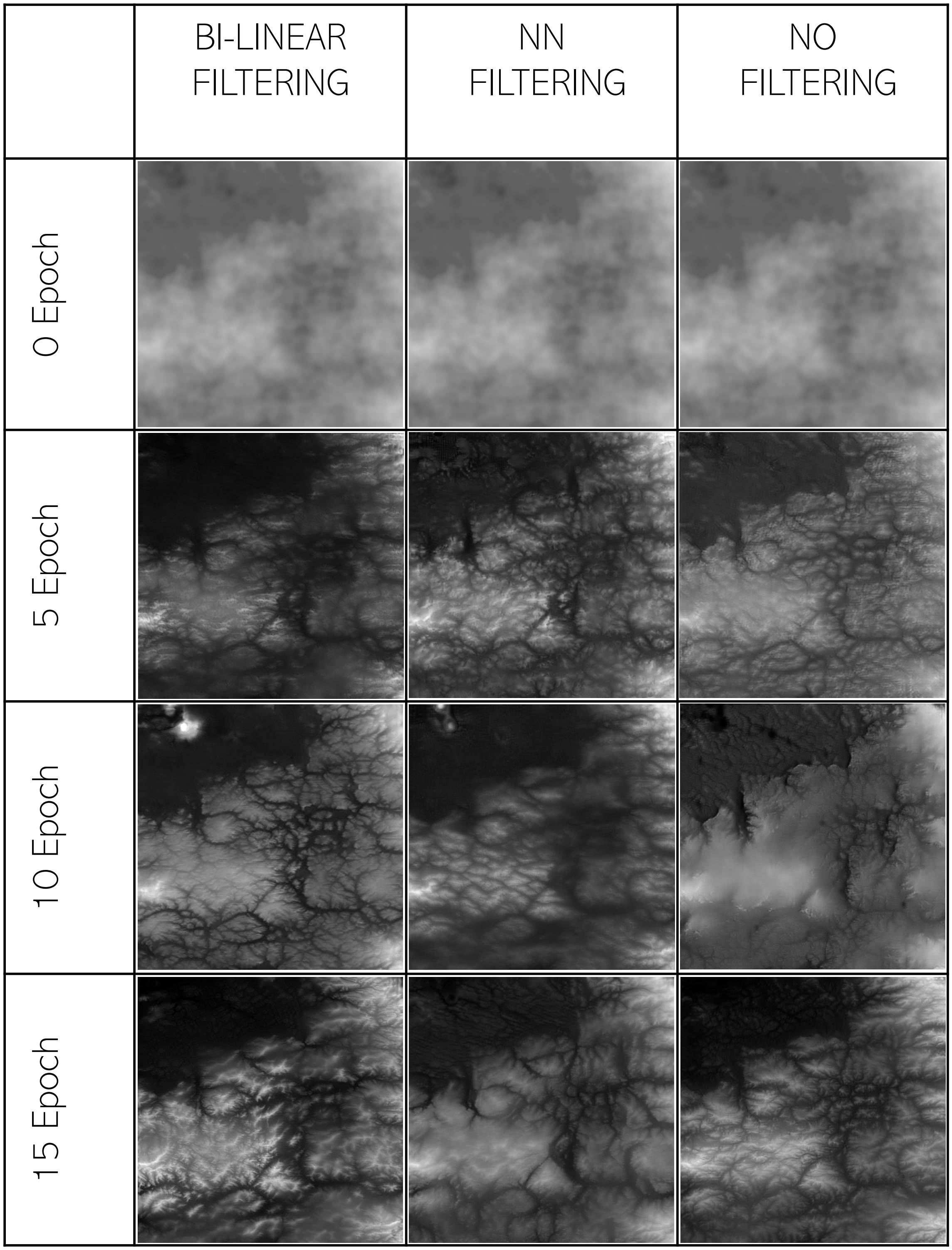


Figure 4.6. The checkerboard artifacts reduces with the training time and the can be mitigated with the resizing of convolution kernels.

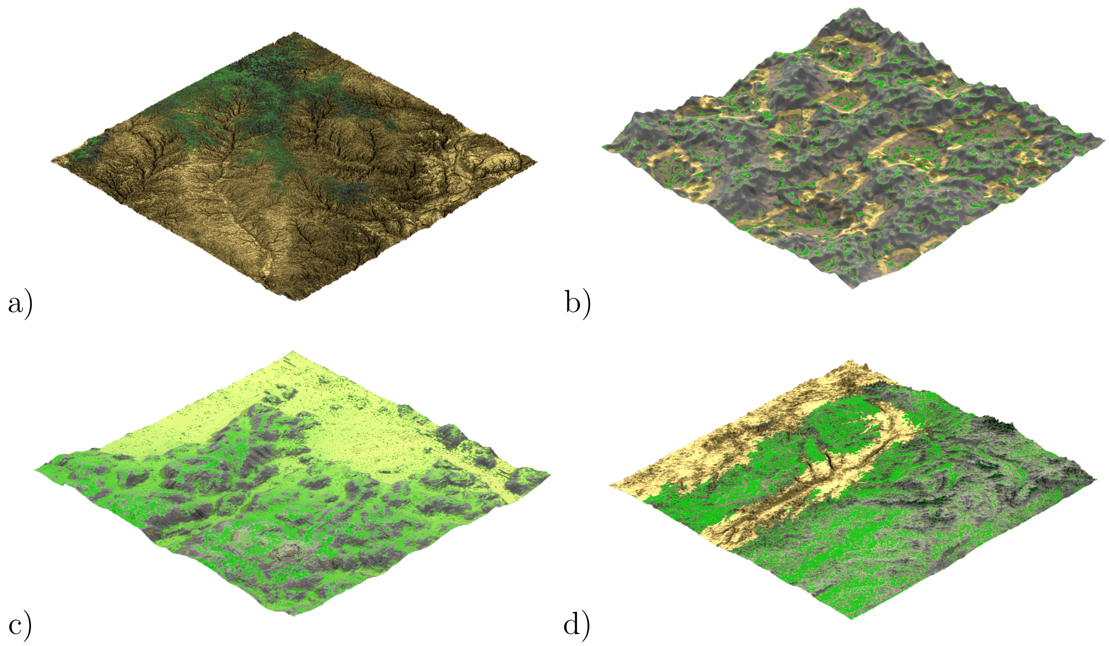


Figure 4.7. Example of feature transfer: a) Real terrain with strong fluvial patterns from Colombian Amazonian forest area (S01 W072) (PTQM=0.67) and b) synthetic terrain generated by thermal erosion (PTQM=0.46). c) Synthetic features transferred to real terrain worsen its perceived visual quality (PTQM=0.49) and d) real features transferred to synthetic terrain improve it (PTQM=0.63).

Table 4.4.

Image pairing for Experiment 2. (Boldface pairs are reused from E1.)

	S	R2S	S2R
R	**R-S**	R2S-R	S2R-R
S	•	R2S-S	S2R-S
R2S	•	•	S2R-R2S

Table 4.5.

Geomorphon value averages for our terrains in Experiments 1 and 2 (Part-I).

CATEGORY	DEPRESSION	SUMMIT	FLAT	VALLEY	RIDGE
RG-AVG:	0.007	0.018	0.166	0.224	0.250
RG-STDEV:	0.004	0.013	0.161	0.058	0.052
RF-AVG:	0.154	0.185	0.124	0.529	0.535
RF-STDEV:	0.217	0.249	0.137	0.246	0.249
RA-AVG:	0.072	0.082	0.158	0.386	0.396
RA-STDEV:	0.156	0.145	0.131	0.165	0.167
RS-AVG:	0.049	0.068	0.371	0.363	0.381
RS-STDEV:	0.051	0.044	0.185	0.117	0.121
RC-AVG:	0.043	0.054	0.286	0.322	0.337
RC-STDEV:	0.036	0.037	0.204	0.120	0.116
SP-AVG:	0.000	0.000	0.198	0.023	0.024
SP-STDEV:	0.000	0.000	0.004	0.001	0.001
SR-AVG:	0.016	0.012	0.000	0.040	0.081
SR-STDEV:	0.002	0.002	0.000	0.004	0.010
SM-AVG:	0.001	0.001	0.178	0.030	0.032
SM-STDEV:	0.001	0.001	0.004	0.002	0.002
ST-AVG:	0.048	0.051	0.000	0.159	0.147
ST-STDEV:	0.003	0.002	0.000	0.005	0.006
SF-AVG:	0.006	0.018	0.001	0.145	0.181
SF-STDEV:	0.002	0.002	0.000	0.006	0.005
SC-AVG:	0.024	0.024	0.001	0.061	0.066
SC-STDEV:	0.002	0.003	0.001	0.004	0.005

Table 4.6.

Geomorphon value averages for our terrains in Experiments 1 and 2 (Part-II).

CATEGORY	HOLLOW	SPUR	SHOULDER	SLOPE	FOOTSLOPE
RG-AVG:	0.388	0.432	0.483	0.551	0.574
RG-STDEV:	0.122	0.122	0.103	0.061	0.109
RF-AVG:	0.460	0.550	0.405	0.396	0.419
RF-STDEV:	0.134	0.158	0.241	0.105	0.222
RA-AVG:	0.408	0.483	0.526	0.431	0.561
RA-STDEV:	0.124	0.142	0.209	0.095	0.203
RS-AVG:	0.336	0.427	0.481	0.314	0.529
RS-STDEV:	0.091	0.109	0.186	0.108	0.194
RC-AVG:	0.343	0.406	0.521	0.397	0.571
RC-STDEV:	0.114	0.129	0.193	0.109	0.187
SP-AVG:	0.135	0.141	0.930	0.587	0.922
SP-STDEV:	0.003	0.003	0.004	0.004	0.005
SR-AVG:	0.581	0.518	0.012	0.974	0.013
SR-STDEV:	0.024	0.017	0.005	0.014	0.005
SM-AVG:	0.148	0.155	0.922	0.596	0.917
SM-STDEV:	0.002	0.003	0.005	0.004	0.005
ST-AVG:	0.766	0.869	0.011	0.777	0.009
ST-STDEV:	0.010	0.012	0.002	0.007	0.002
SF-AVG:	0.972	0.955	0.025	0.694	0.023
SF-STDEV:	0.014	0.017	0.004	0.009	0.004
SC-AVG:	0.642	0.668	0.017	0.914	0.016
SC-STDEV:	0.017	0.018	0.005	0.010	0.003

CHAPTER 5. RESULTS

An example in Figure 5.1 shows a procedural terrain and the distribution of its landforms based on geomorphons as well as a real terrain with its accompanying features. The feature vector of the geomorphons is organized so that the visually plausible ones are on the right hand side. The real terrain was ranked as highly visually plausible (77%) in our user study and the procedural terrain was on the opposite scale (49%) as can be also seen in the distribution of the geomorphons. We then used deep learning to transfer the features from the procedural to real and vice versa and we show the corresponding distribution of the geographics features that indicates that the distributions of the geomorphons changed so that the high-ranked worsen and low-ranked improved. This quantitative validation has been then

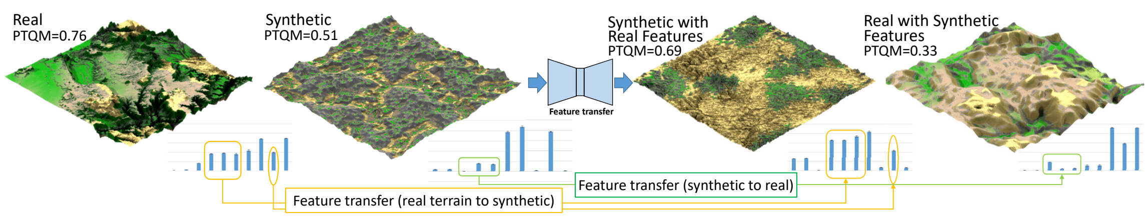


Figure 5.1. The real terrain from the state of Arizona in the USA with complex geomorphological patterns has PTQM=0.76 of top visually plausible and it ranked as top 78% in our perceptual study. The synthetic terrain models with patterns generated by thermal erosion is has PTQM=0.51 and it ranked as 49% in the study. The corresponding geomorphons show the distribution of patterns in each model with strong presence of valleys, ridges, and hollows landform in real terrain that were not so present in the synthetic variety. By using a CycleGAN, we transferred the visually important features to the procedural terrain (orange arrows) and we transferred the features in synthetic terrain to the real terrain (green arrow). The second perceptual study showed that the transferred features improved to PTQM=69 (77% ranking in our study) and transferring the visually unattractive features from procedural terrain to real demoted its PTQM=0.33 (29%). The transferred features are circled in the corresponding graphs of geomorphons.

confirmed by the second user study that showed that the procedural terrain after the style transfer improves its visual plausibility by to 69% and the real terrain worsens to 29%.

Below we describe results of our two experiments and feature transfer. We show results of E1 and E2, discuss the features in geomorphons and the feature transfer. Finally, we introduce the perceptual terrain quality metric PTQM (see also the supplemental material).

5.1 Perceptual Experiments

Experiment 1: The first Mechanical Turk experiment assigned each image a number of how many times it was selected as more realistic. We normalized the counts so that the best image had a score of 1.0. We then calculated the average, standard deviation, mean, and range for each category of R and S from Table 4.1. The sorted results by the average value are shown in Figure 5.3 top. The ranking of terrains from least visually plausible to the best was: SR-SC-SF-SM-SP-ST-RS-RC-RA-RF-RG. All synthetic terrains were perceived as visually worse as compared to the real ones. The most visually plausible synthetic terrains are generated by thermal erosion. See also Table 5.4 for the actual numbers.

We have also calculated the average and standard deviation of values of ranking of all images in the sets S and R. An unpaired T-Test evaluation suggested that the difference is statistically significant with the two-tailed $p - value < 0.01$, $DF = 283, t = 17.91$ & $\alpha = 0.01$.

Experiment 2: The second experiment repeated E1 with the addition of pairs of images with transferred features. Our assumption was that the features transferred from real terrains to synthetic would improve their ranking and that the transfer of features from synthetic to real terrains will worsen their ranking.

Similarly to E1 we have normalized the rank of each image and calculated statistics for each category. Figure 5.3 shows the result and Table 5.4 shows the actual numbers.

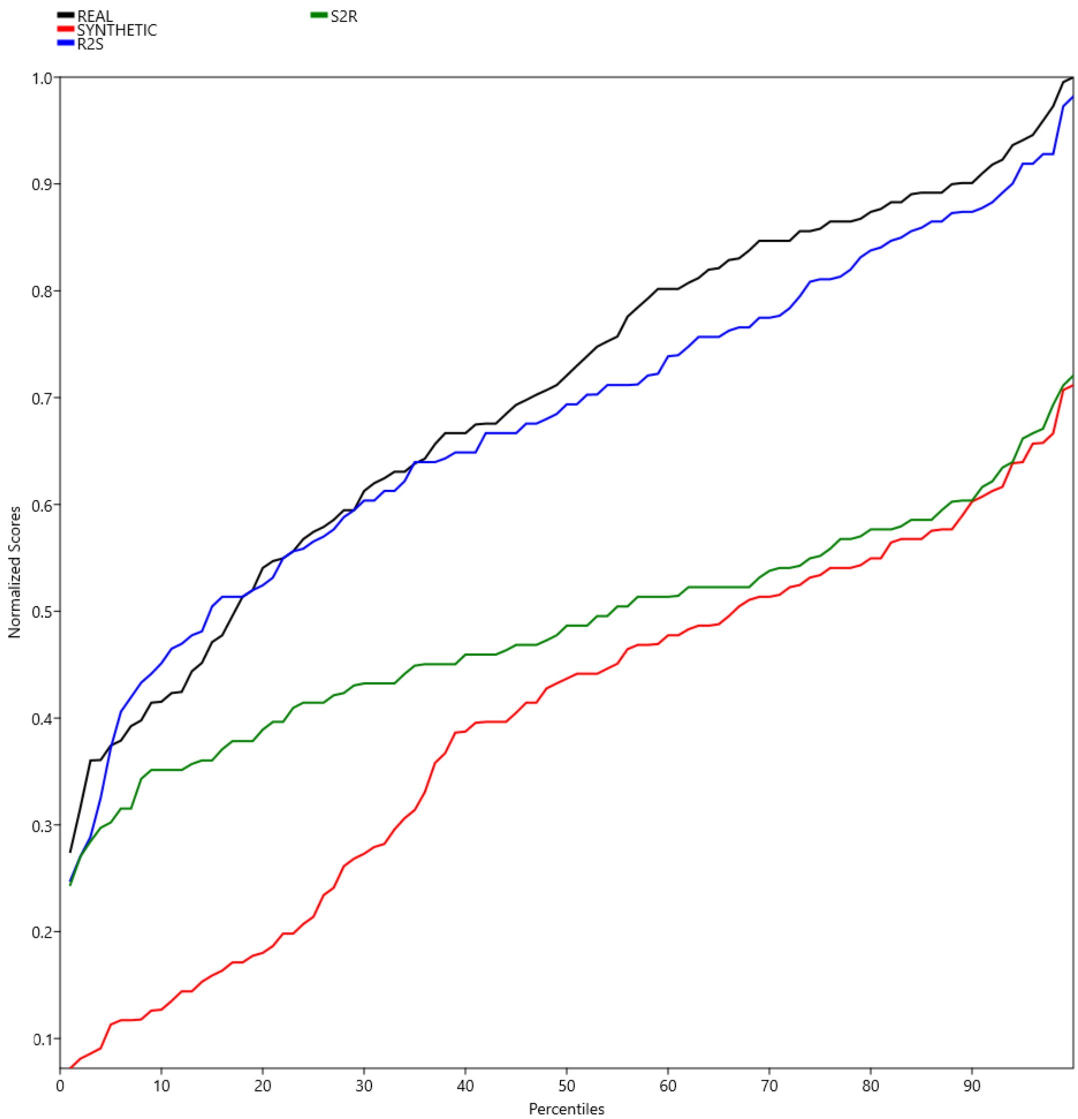


Figure 5.2. Percentile Plot of Normalized Scores from Experiment-2.

Simply stated, the perceptual experiment suggests that synthetic terrains in our data-set are perceived as visually significantly worse than the real ones.

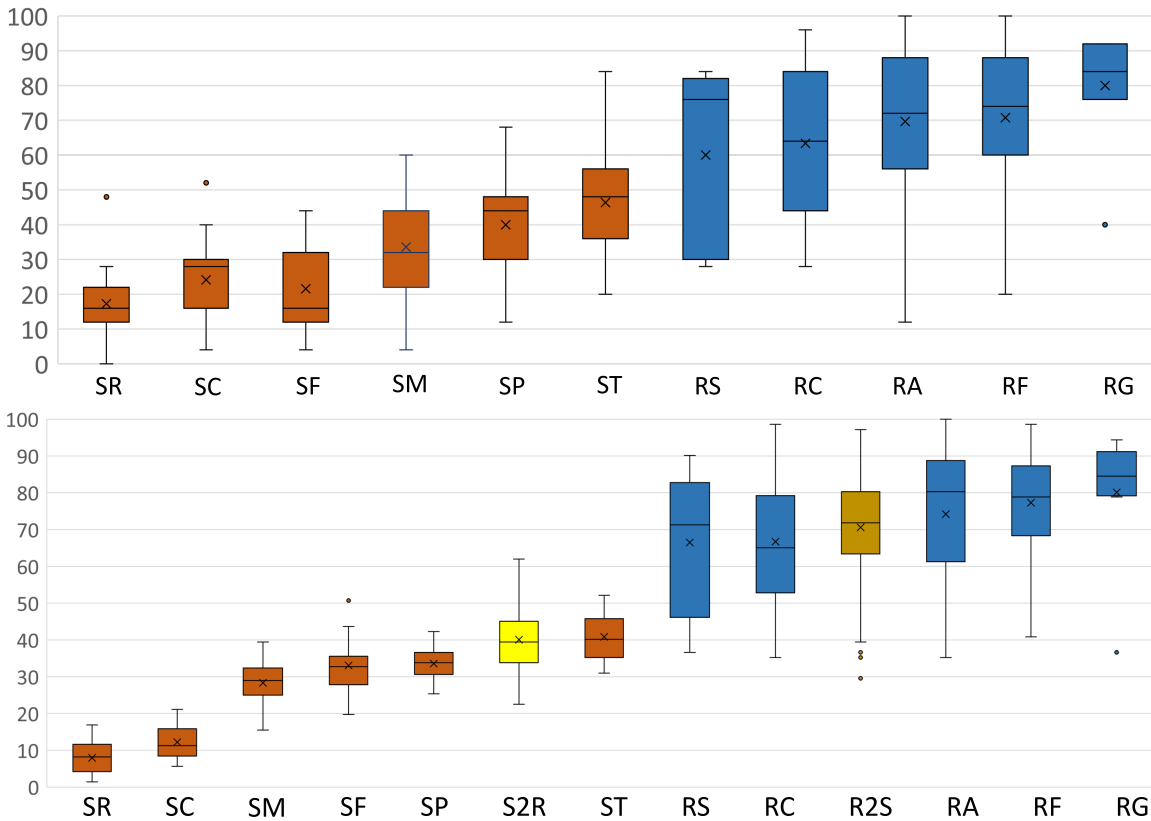


Figure 5.3. Distribution of perception scores from experiments: E1 and E2. Perceptual ranking of terrains from E1 (top) and E2 (bottom). The abbreviations are from Table 4.1 and the terrains are sorted by the average perceived visual quality from worse (left) to the best (right). While the order of the rankings in E2 is very similar to E1, note that the S2R synthetic terrains improved with features from real terrains ranked high. At the same time, real terrain with features transferred from procedural R2S ranked lower. The figure has been plotted based on their average scores. The $\times$, $\bullet$, and the $-$ sign represent the mean, outlier points, and the median markers respectively. These plots should be interpreted with care as they should not be used for interpret or infer statistical significance of the difference among terrain categories.

The order of terrain categories is exactly the same as in E1 that confirms the validity of both tests. The categories with transferred features ranked as expected. The synthetic terrains enhanced with features from real terrains R2S ranked 10th, which is better than some of the real terrains, but better than all synthetic ones. This confirms our initial hypothesis that feature transfer has an important effect on

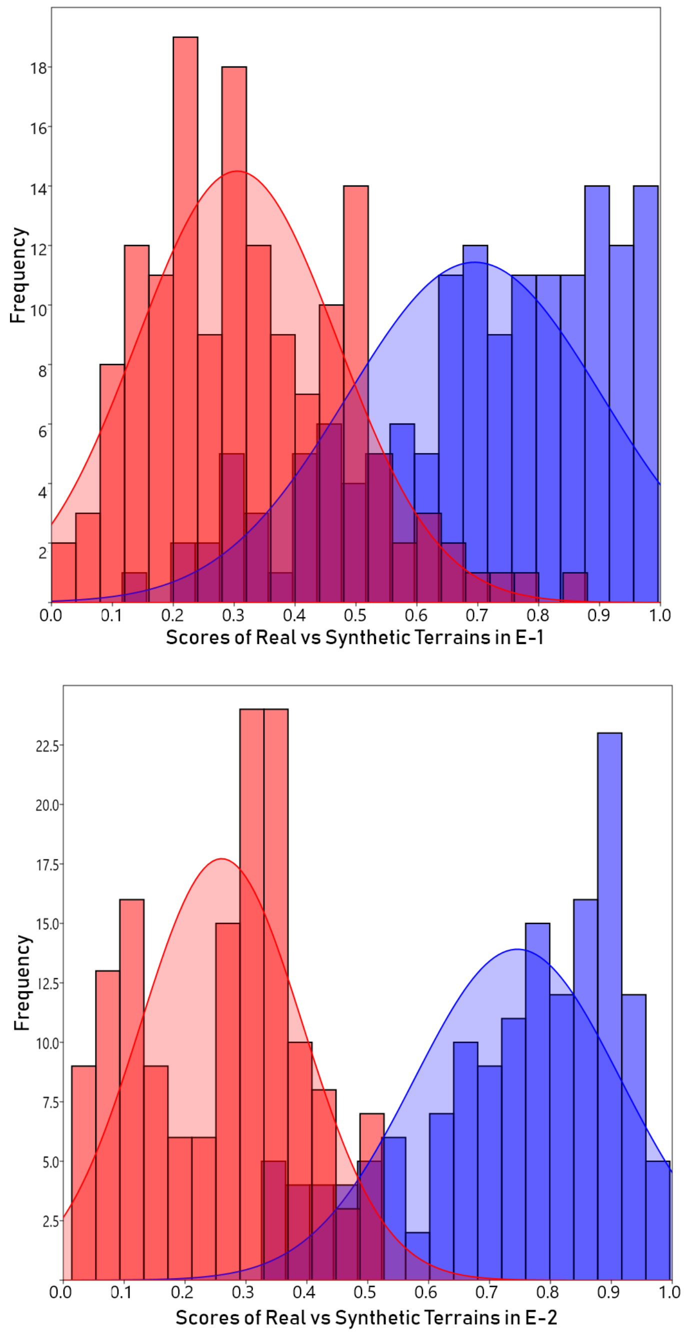


Figure 5.4. Frequency Distribution comparison of our normalized scores from Experiments E-1 and E-2 respectively. The match shows the robustness and validity of our method.

Table 5.1.

The average perceptual scores of all our terrain samples in each test category.

Type	**Category**	**E1**	**E2**
Real (R)	RG	0.80	0.80
	RF	0.71	0.77
	RA	0.70	0.74
	RC	0.63	0.67
	RS	0.60	0.66
Synthetic (S)	ST	0.46	0.41
	SP	0.40	0.34
	SM	0.34	0.28
	SF	0.22	0.33
	SC	0.24	0.12
	SR	0.17	0.08
Transfers (T)	R2S	N/A	0.71
	S2R	N/A	0.40

terrain perception. Similarly, the real terrain with transferred procedural features S2R ranked significantly worse than all real terrain and even worse than thermal erosion simulation at 6th place. This again confirmed our hypothesis that features of synthetic terrains do not contribute to visual quality of terrains.

5.2 Statistical Tests

We performed statistical tests on our normalized perceptual scores to determine if there are any differences in perception of our terrain data groups:R, S, R2S and S2R. We state the null hypothesis, H_0 for our 6 statistical tests in E2 as follows: *"There are no significant differences in the visual perception scores between our terrain data groups."*. We used t-Tests to compare the means and variances of the perception scores. We have summarized the results for our statistical tests in Table 4. For testing our candidates in E2, we used the significance level of $\alpha = 0.01$, and get the statistics for, R versus R2S ($p - value = 0.02$, $DF = 149, t = 2.26$), R versus S2R ($p-value < 0.01$, $DF = 149, t = 22.10$), R versus S ($p-value < 0.01$, $DF = 149, t = 22.59$), R2S versus S2R ($p-value < 0.01$, $DF = 149, t = -23.52$), R2S versus S ($p-value < 0.01$, $DF = 149, t = 29.12$) and S2R versus S ($p-value < 0.01$, $DF = 149, t = 10.79$). It is evident from the tests (Table 5.2) that the perception scores are statistically different between the terrain groups meaning that our observers perceived and like the terrains on different scales except the candidate R v R2S signifying that there are features in real terrains that increase the visual plausibility. Therefore, we can safely reject our null hypothesis for R v R2S and state that there is a significant difference in perception of Real Terrains (R), Synthetic Terrains (S), Synthetic Terrains with Real features (R2S) and Real Terrains with Synthetic features (S2R).

We initially performed an ANOVA (E1: $calculated\ F-Value = 320.91$, $critical\ F-value = 3.87$, $p < 0.01$, $df = 298$) (E2: $calculated\ F-Value = 465.78$, $critical\ F-value = 2.61$, $p < 0.01$, $df = 596$) to determine if there are any significant difference in the variances of the scores and after establishing that there are differences in the groups, we proceeded with the t-tests to determine among which groups the significant differences lie. Additionally, a post-hoc test (Tukey's HSD (honestly significant difference) test) indicated that there is no statistically significant difference in the perception scores between the terrain groups, R and R2S with a $p-value = 0.0511$ and standard error of 1.0938 while there is a statistically significant difference between the rest of the terrain groups with $p-value < 0.001$ and standard error of 1.0938 with $\alpha = 0.01$ which is consistent with t-Test results.

Table 5.2.

The table shows the statistical significance of each terrain set compared with the other set from our experiments: E1 and E2. The ✓ implies that the terrain set in the vertical column are statistically significant than the terrain set in the horizontal row, × to suggest that the difference is not statistically significant and • to suggest that the test is not available or compared already.

	R	S	R2S	S2R
R	•	✓	×	✓
S	•	•	✓	✓
R2S	•	•	•	✓
S2R	•	•	•	•

Table 5.3.

The Average (AVG), Median (MED), Mode (MODE), RANGE (RNG), Standard Deviation (STDEV), Standard Error (SE), and 95% Confidence Interval (95% C.I.) of the normalized scores for the terrain set: E1.

		E2						
T	**Ab.**	**AVG**	**MED**	**MODE**	**RNG**	**STDEV**	**SE**	**95% C.I.**
R	RG	80	84	92	52	19	7	14
	RF	71	74	88	80	20	3	5
	RA	70	72	92	88	22	3	6
	RC	63	64	96	68	21	5	9
	RS	60	76	N/A	56	28	12	24
S	ST	46	48	48	64	17	3	7
	SP	40	44	48	56	13	3	5
	SM	34	32	36	56	14	3	6
	SF	22	16	16	40	12	2	5
	SC	24	28	28	48	11	2	4
	SR	17	16	20	48	10	2	4
2	R2S	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S2R	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 5.4.

The Average (AVG), Median (MED), Mode (MODE), RANGE (RNG), Standard Deviation (STDEV), Standard Error (SE), and 95% Confidence Interval (95% C.I.) of the normalized scores for the terrain set: E2.

		E2						
T	**Ab.**	**AVG**	**MED**	**MODE**	**RNG**	**STDEV**	**SE**	**95% C.I.**
R	RG	80	851	N/A	58	19	7	13
	RF	77	79	86	58	13	2	3
	RA	74	80	89	65	19	3	5
	RC	67	65	63	63	17	4	8
	RS	67	71	N/A	53	20	8	16
S	ST	41	40	32	21	7	1	3
	SP	34	34	34	17	5	1	2
	SM	28	29	32	24	6	1	2
	SF	33	33	35	31	8	2	3
	SC	12	11	10	16	5	1	2
	SR	8	8	13	16	4	1	2
2	R2S	71	72	70	68	13	2	2
	S2R	40	39	33	39	9	1	1

5.3 Thurstone Scaling

We constructed and utilized a pairwise comparison strategy in our experiments as it is efficient in capturing subjective judgements from our participants in terms of identifying the differences in perception of terrains. Because of the nature of our testing platform (Mechanical Turk), we do not have control over the balance of the study as the respondents can decide to not vote for all the image pairs in an experiment. Additionally, we were not able to construct a complete design as in the terrains of a subgroup are not compared with all of the other terrains of another. We achieve balance in our study by forcing each image pair will be voted by at least five different and unique participants.

For efficiently capturing both the ranking and the magnitude of differences between these conditions in a study that has a partially balanced incomplete design such as ours, a scaling method can be utilized. Perez-Ortiz and Mantiuk (2017) suggested a pairwise comparison scaling method that is based on Thurstone Case V model to measure the attitude of respondents based on probabilities (5.5. Therefore, to perform such a scaling, we constructed a comparison matrix for responses from each of the observers based on the conditions being compared in our experiments E1 and E2 namely, R, S, R2S and S2R. The comparison matrix can then be utilized to construct standardized scores and probabilities based on a observer model such as Thurstone Case V model.

As a preliminary step before standardizing the scores, we performed outlier analysis to look for any potential outliers. One such potential outlier is show in Figure 5.5. After carefully investigating all the potential outliers based on a inter-quartile normalised score threshold, we couldn't justify removing respondent observations from the comparison matrix because of the nature of our study design and our survey platform (mTurk) in which some respondents may choose to answer the complete set of comparison pairs whereas others may only choose to do a few. Therefore, we did not remove the remove the responses from the dataset that are

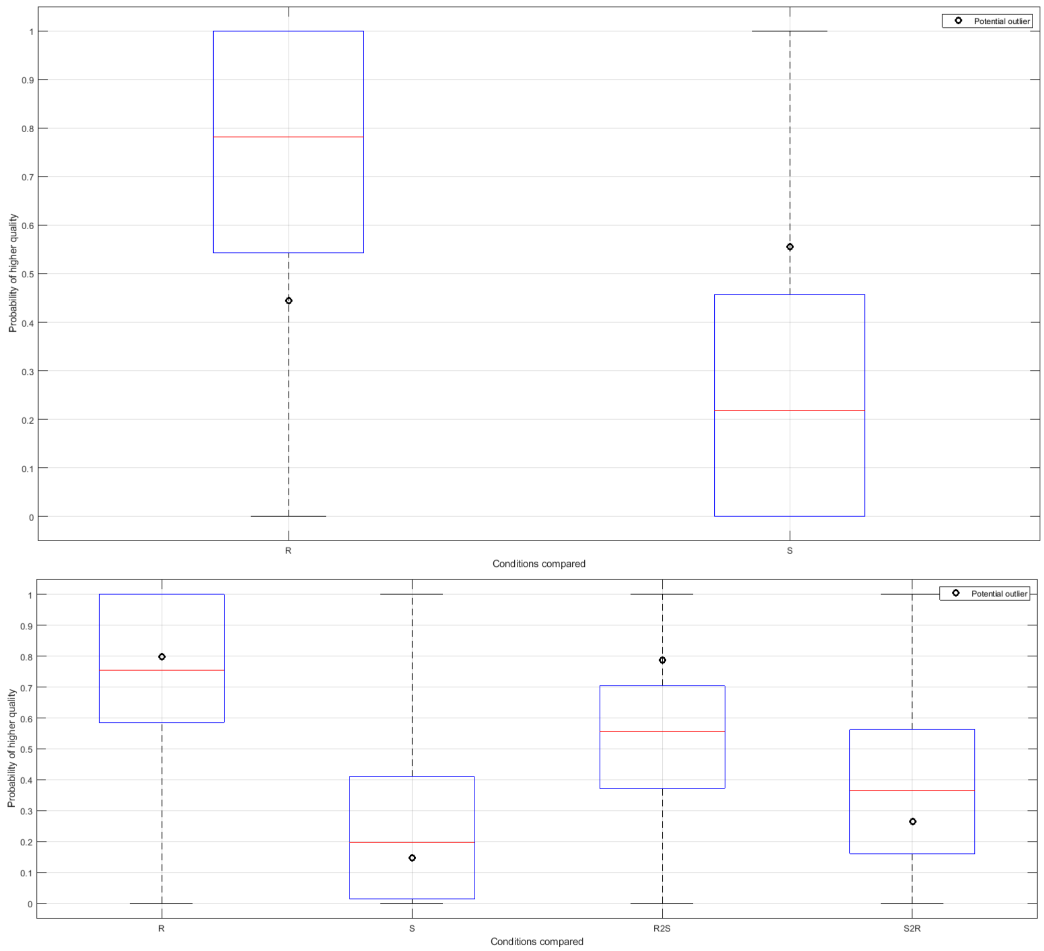


Figure 5.5. Distribution of probabilities of selection of one condition over the other for general perceived quality from experiments: E1 (top) and E2 (bottom). The black circle sign represent the answers of a potential outlier.

shown as potential outliers after careful analysis and investigation determining that they chose to answer more from the comparison set when compared to the rest of the respondents (Perez-Ortiz & Mantiuk, 2017). We then proceeded with standardization of scores based on Thurstonian scaling.

We performed an ANOVA on both the experiments E1 and E2 for identifying if there are any significant differences in standardized perception scores for our terrain categories with the null hypothesis, H_0 for our statistical test in our experiments, E1 and E2 as follows: *"There are no significant differences in the standardized perception scores scores between terrain categories."*. For E1, the

ANOVA test yielded the following statistic: ($critical\ F - value = 3.8507$, $Between\ Groups\ DF = 1$, $Within\ Groups\ DF = 998$, $calculated\ F - value = 3336.037$, $p - value < 0.001$) and For E2, the ANOVA test yielded: ($Critical\ F - value = 2.6093$, $Between\ Groups\ DF = 3$, $Within\ Groups\ DF = 1996$, $Calculated\ F - value = 2592.893$, $p - value < 0.001$). Therefore, for both the experiments we can reject the null hypothesis and conclude that there are significant differences in perception scores by terrain categories: R, S, R2S and S2R. Additionally, a post-hoc test (Tukey's HSD test) also indicated that there is statistically significant difference in the scaled scores based on Thurstonian scaling between all the terrain groups with a $p - value < 0.001$ and a standard error of 0.0077 with $\alpha = 0.01$ which is consistent with our t-Test results.

For further establishing where the difference in our terrain candidates lie, we performed a Two-Sample t-Test Assuming Unequal Variances for testing our candidates in E2. We used the significance level of $\alpha = 0.05$, and get the statistics for, R versus S ($p - value = 0$, $DF = 983, t = 84.75$), R versus R2S ($p - value < 0.01$, $DF = 987, t = 9.67$), R versus S2R ($p - value < 0.01$, $DF = 908, t = 46.20$), R2S versus S2R ($p - value < 0.01$, $DF = 954, t = 36.28$), R2S versus S ($p - value = 0$, $DF = 998, t = 71.54$) and S2R versus S ($p - value < 0.01$, $DF = 960, t = 27.45$). We have summarized the same results in the table that follows: Table 5.5.

It is evident from the multiple statistical tests that the perception scores are statistically different between the terrain groups meaning that our observers perceived and like the terrains on different scales with the following ranking: Real Terrains (R), Synthetic Terrains with Real features (R2S) and Real Terrains with Synthetic features (S2R), Synthetic Terrains (S). The ranking has been visualized in the following plots (Figure 5.6 and Figure 5.7).

Table 5.5.

The table shows the statistical significance between pair of standardized scores from our terrain sets compared from Experiment E2 that is based on thurstonian scaling. The ✓ implies that the condition pair in vertical column is statistically significant than the terrain set in the horizontal row, × to suggest that the difference is not statistically significant and • to suggest that the test is not available or compared already.

	R	S	R2S	S2R
R	•	✓	✓	✓
S	•	•	✓	✓
R2S	•	•	•	✓
S2R	•	•	•	•

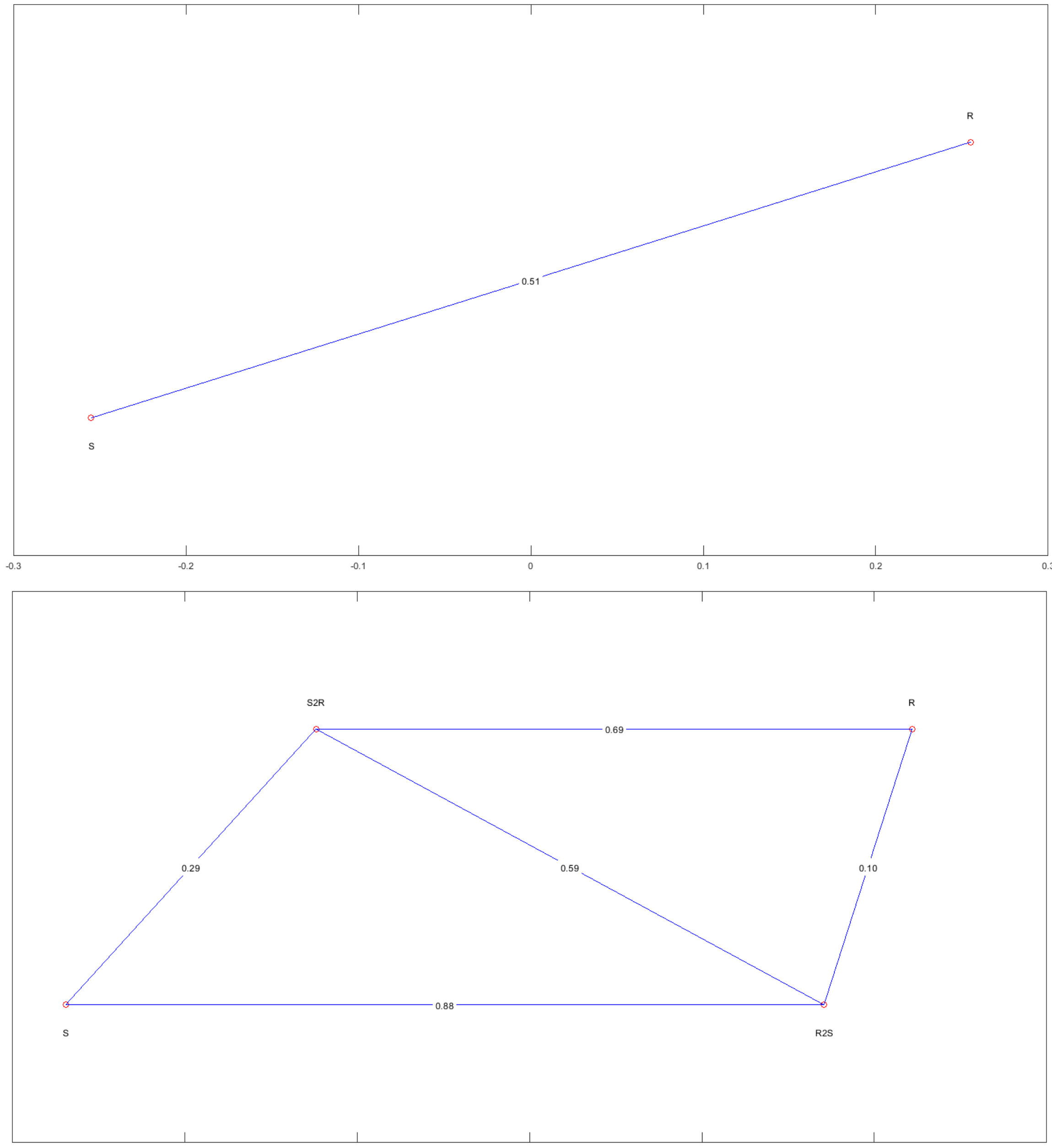


Figure 5.6. Triangle Plot of general perceived quality from experiments: E1 (top) and E2 (bottom) for interpretation of data in terms of statistical significance with 95% confidence. The categories in the triangle plot connected with continuous lines indicate statistical significance whereas the categories connected with dashed lines indicate a lack of evidence for establishing statistical significance. In our case, all the terrain categories are different from each other with statistical significance and hence there are no dashed lines.

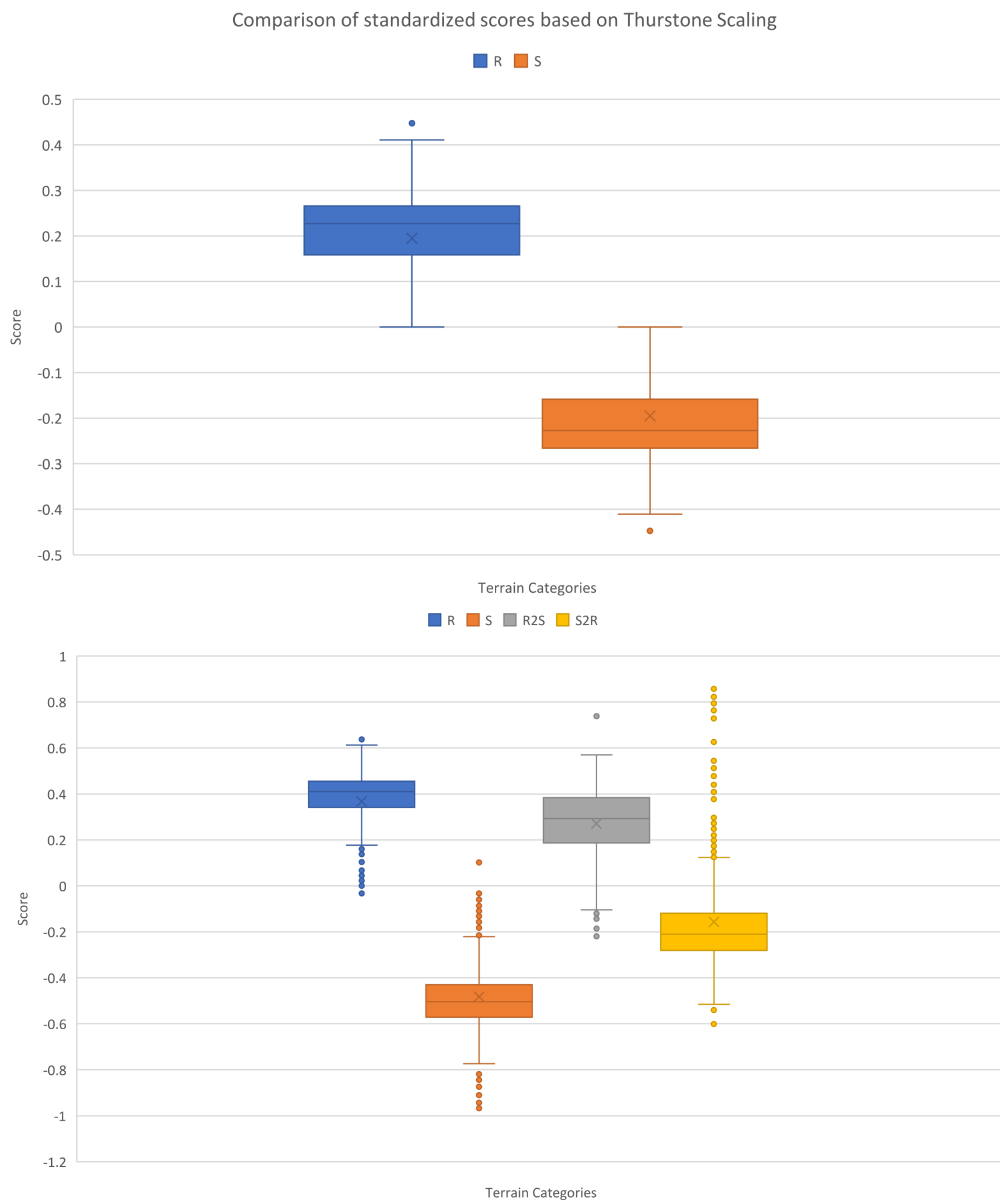


Figure 5.7. Distribution of standardized scores from experiments: E1 (top) and E2 (bottom). The $\times$, $\bullet$, and the $-$ sign represent the mean, outlier points, and the median markers respectively.

5.4 Just Objectionable Differences (JOD)

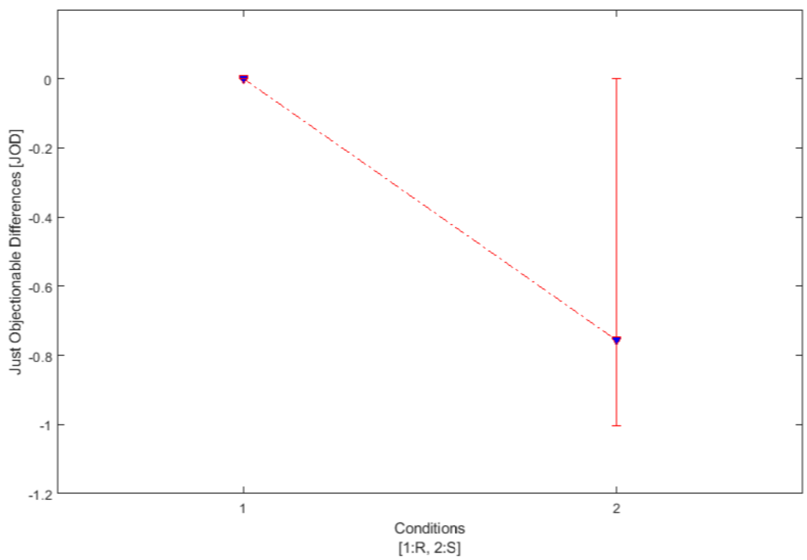
The standardized scores based on Thurstonian scaling mechanism are then converted to distances for the underlying quality scores to determine Just Objectional Differences [JODs]. We made use of publicly available MATLAB code by (Perez-Ortiz & Mantiuk, 2017) with a few modifications to fit our data to determine the JOD scores with bootstraping upto 500 samples as suggested. As we have already identified Real Terrain (R) as the top category from our terrain groups, we can set the condition as our anchor point (reference condition) for JOD, as it is a relative scoring scheme. A visualization of these scaling results are shown in the figure 5.8.

We cannot directly compare the JOD scaling values between terrain subgroups as these values are interlinked and interdependent on each other. Based on testing strategies for JOD comparison scales, we performed statistical tests to determine differences between two conditions based on the covariance matrix and JOD scores on our scaled data set to identify that the JOD score difference between two conditions is 0 (Perez-Ortiz & Mantiuk, 2017). The captured covariance matrix for the pairwise comparison of our conditions [R, S] in E1 is given by the matrix, Σ_{E1} :

$$\Sigma_{\mathrm{E1}} = \begin{bmatrix} 0 & 0 \\ 0 & 0.1024 \end{bmatrix}$$

We state the null hypothesis, H_0 for our statistical test in our experiment, E1 as follows: *"There are no significant differences in the JOD scores between conditions R and S."*. We utilized a two-tailed F-test to test our hypothesis, with a significance level of $\alpha = 0.025$, and get the statistics for, R versus S ($critical\ F-value = 1.1920$, $Numerator\ DF = 499$, $Denominator\ DF = 499$, $calculated\ F-value = 7.3828$).

The covariance matrix for the pairwise comparison of our conditions [R, S, R2S, S2R] in E2 is given by the matrix, Σ_{E2}:



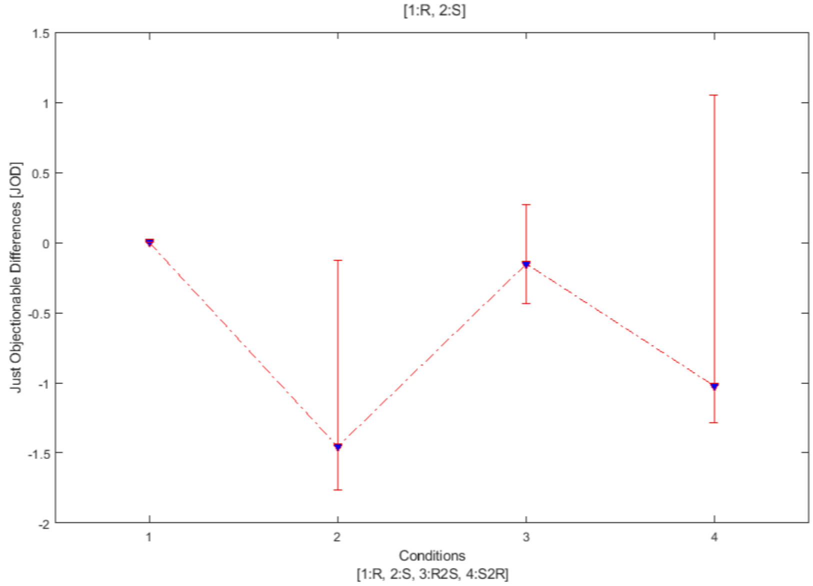


Figure 5.8. The graph shown above visualizes the scaling results for experiments E1 and E2 along with their 95% confidence intervals for our dataset. The first condition (R) is our reference condition therefore, it is always set at 0 and hence there are no confidence intervals. A difference of 1 JOD unit indicates that 75% of the participants chose a condition over the other.

$$\Sigma_{\mathrm{E2}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.1763 & 0.0033 & 0.1337 \\ 0 & 0.0033 & 0.0401 & 0.0055 \\ 0 & 0.1337 & 0.0055 & 0.4356 \end{bmatrix}$$

Similar to E1, We state the null hypothesis, H_0 for our 6 statistical tests in our experiments, E2 as follows: *"There are no significant differences in the JOD scores between conditions."*. Likewise, we utilized a two-tailed F-test to test our hypothesis, with a significance level of $\alpha = 0.025$, and get the statistics for, R versus S ($critical\ F-value = 1.1920$, $Numerator\ DF = 499$, $Denominator\ DF = 499$, $calculated\ F-value = 8.2558$), R versus R2S ($critical\ F-value = 1.1920$, $Numerator\ DF = 499$, $Denominator\ DF = 499$, $calculated\ F-value = 3.7805$), R versus S2R ($critical\ F-value = 1.1920$, $Numerator\ DF = 499$, $Denominator\ DF = 499$, $calculated\ F-value = 2.3530$), S versus R2S ($critical\ F-value = 1.1920$, $Numerator\ DF = 499$, $Denominator\ DF = 499$, $calculated\ F-value = 6.2144$), S versus S2R ($critical\ F-value = 1.1920$, $Numerator\ DF = 499$, $Denominator\ DF = 499$, $calculated\ F-value = 1.2496$), R2S versus S2R ($critical\ F-value = 1.1920$, $Numerator\ DF = 499$, $Denominator\ DF = 499$, $calculated\ F-value = 1.8794$).

It is evident from the tests (Table 5.6) that the JOD scores are statistically different between the conditions representing our terrain sets: R, S, R2S and S2R in both of our experiments E1 and E2. Therefore, we can reject the null hypothesis stated above and conclude from the results that the respondents view the difference between terrain groups with statistical significance. The results are consistent from our previous findings except the R and R2S terrain comparison category.

Table 5.6.

The table shows the statistical significance between pair of conditions from our terrain sets compared from Experiment E2. The ✓ implies that the condition pair in vertical column is statistically significant than the terrain set in the horizontal row, × to suggest that the difference is not statistically significant and • to suggest that the test is not available or compared already.

	R	S	R2S	S2R
R	•	✓	✓	✓
S	•	•	✓	✓
R2S	•	•	•	✓
S2R	•	•	•	•

5.5 Geomorphons

Each geomorphon can be thought of as a feature vector in 10D space and their spatial distribution can bring further insight into the features and the corresponding data-sets. In Figure 5.9 we show the points corresponding to all our data-sets (R, S, R2S, and S2R) projected from 10D space to 2D by using Principal Component Analysis Wold, Esbensen, and Geladi (1987) that preserves distance among points across the dimensions. Synthetic images are clustered close to each other, while features of real terrains are scattered over a wide area. This is confirmed by the variance of the features as can be seen in graphs in Figure 5.11. When the real features are transferred to synthetic terrains, they tend to scatter the images apart and when synthetic features are transferred to real terrains they tend to get close to each other. This seems to indicate that a high variability in geomorphological features is beneficial for visual plausibility.

Moreover, we visualize domain-wise comparisons among R, R2S, S, and S2R on the distributions of the element-wise geomorphon feature of terrains in Figure 5.10. The geomorphon features of real terrains (blue curve) tend to distribute normally with a wide span. However, the synthetic features (green curve) show significant differences from the real with multi-modal and low-variability

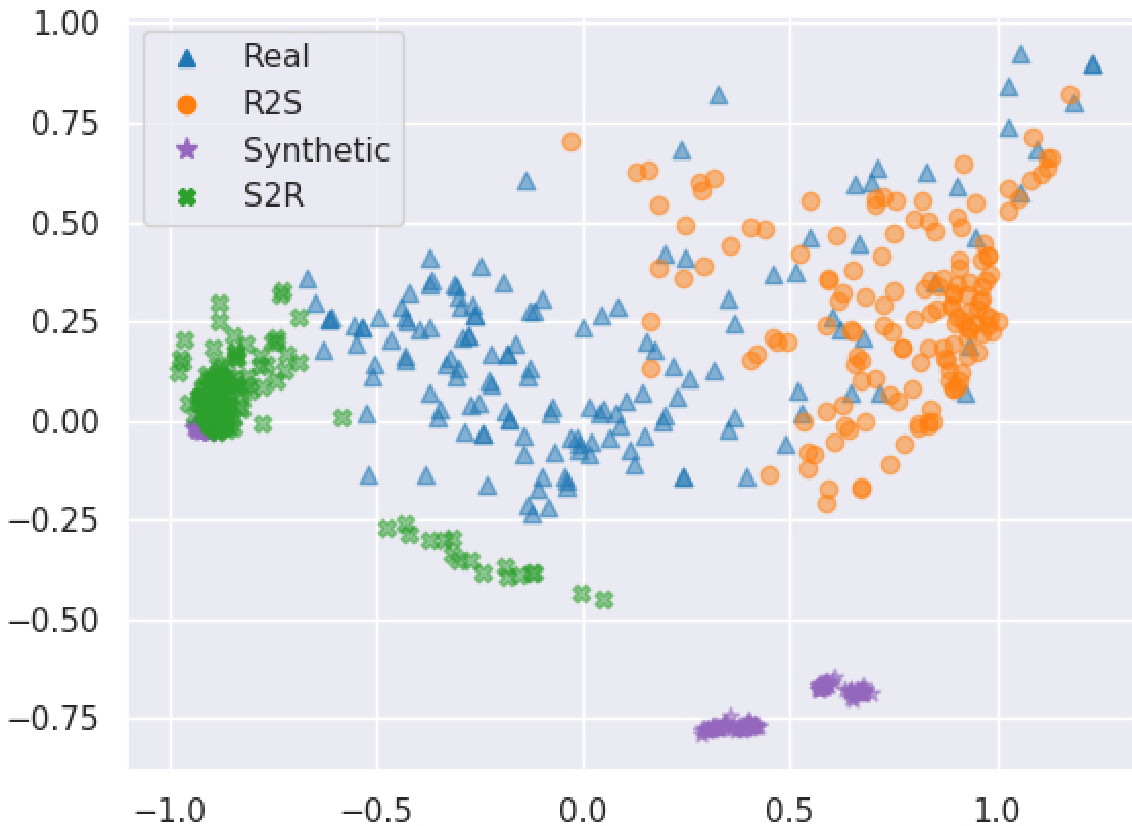


Figure 5.9. Projection of geomorphons from all terrains to 2D. Synthetic terrains are clustered, while real terrains are scattered. Transfer of real features scatters the terrains and transfer of procedural features cluster the resulting terrains. The axes indicate the two projected principal components.

distributions on depression, summit, flat, valley, and ridge (Figure 5.10 top row). We believe the high-peak distributions of synthetic terrains lead to less attractive perceptions than the real. The process of R2S transfer (orange curve) smooths and normalizes the multi-modal high-peak distributions in the synthetic terrains, and improves the perception (refer to Section 5.1). One outlier feature of R2S is the flat geomorphon, which has an extremely narrow distribution on small values near zero, meaning that R2S terrains have very few "flat" elements compared to the real ones. On the other hand, S2R (red curve) contains the low-variability distributed features transferred from the synthetic to the real. However, S2R does not perform a highly fit for the synthetic curves in hollow, spur, shoulder, and footslope. From the visualizations, it seems that lack of geomorphon diversity or variability of individual geomorphon feature in distribution may lead to worse visual plausibility of the terrain.

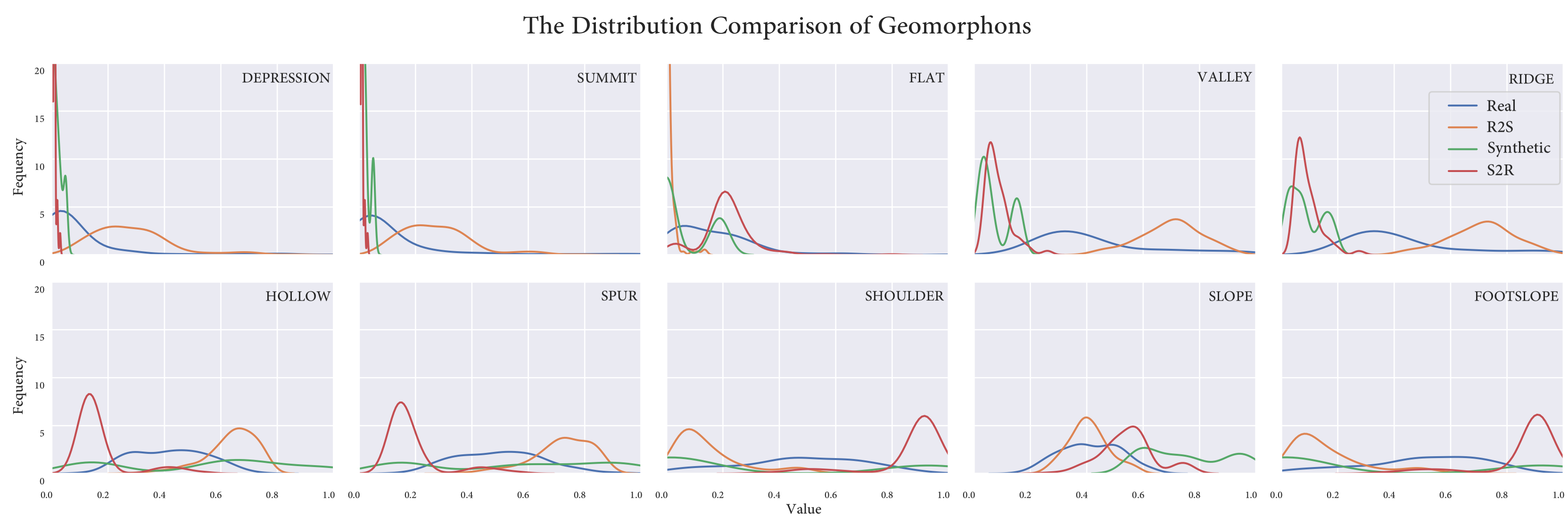


Figure 5.10. The geomorphon feature comparisons among Real, R2S, Synthetic, and S2R.

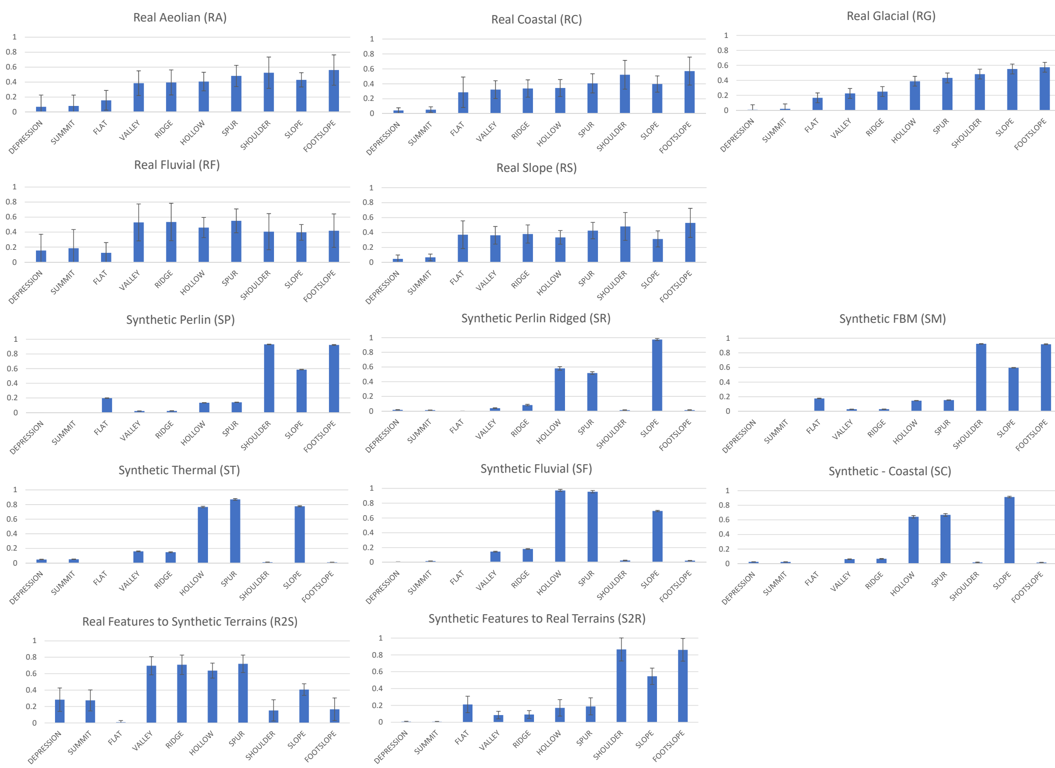


Figure 5.11. Distribution of the detected geomorphons in real and synthetic terrains from our dataset.

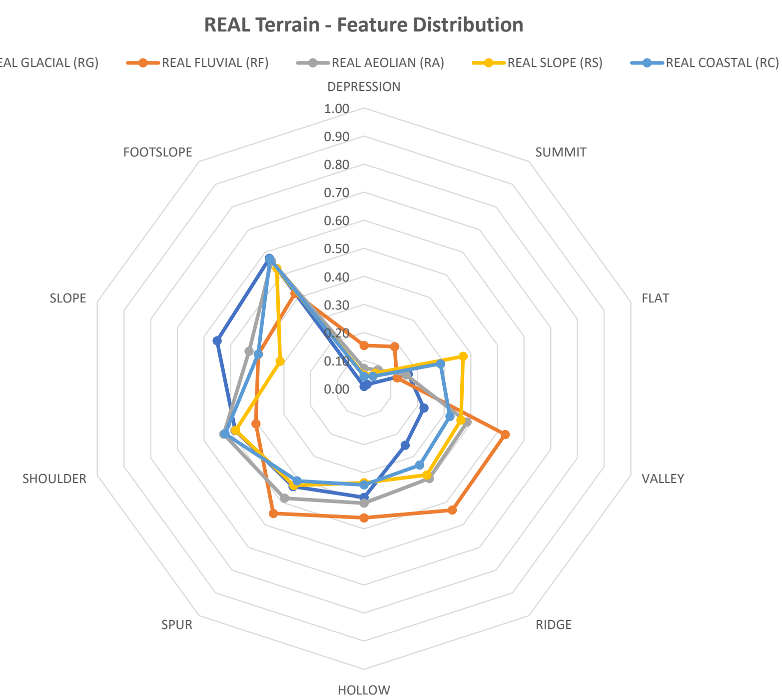


Figure 5.12. The geomorphon feature distribution in Real Terrains.

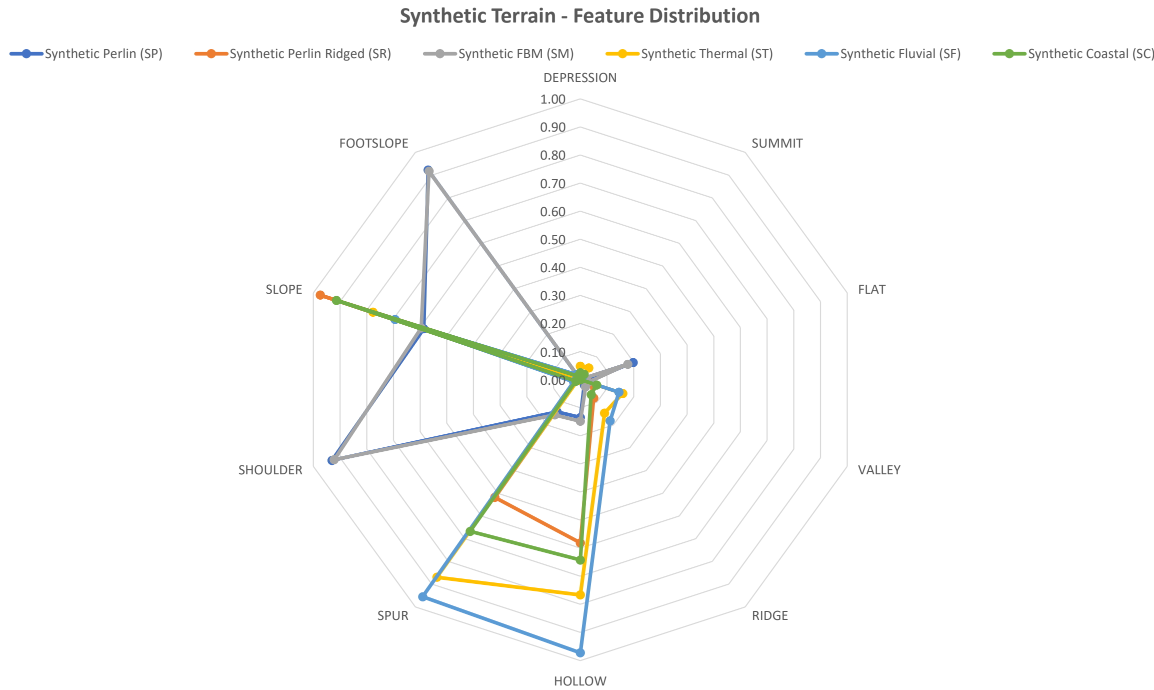


Figure 5.13. The geomorphon feature distribution in Synthetic Terrains.

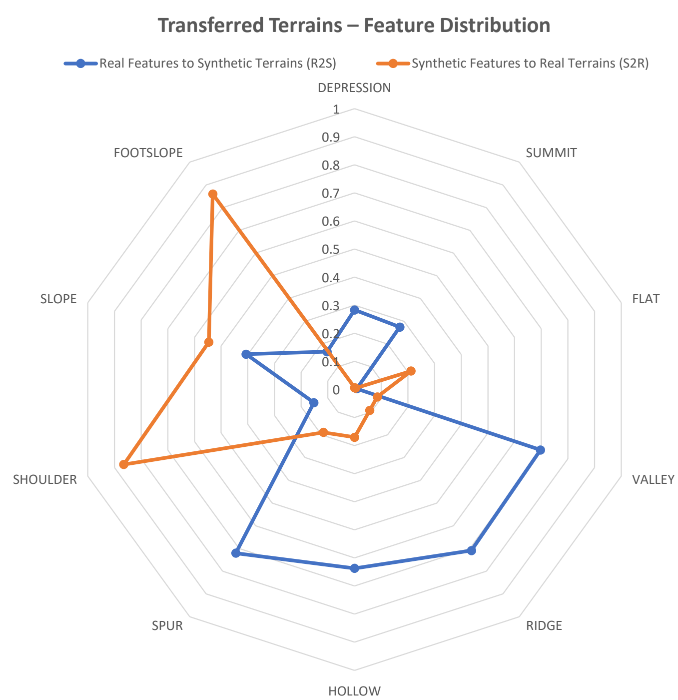


Figure 5.14. The geomorphon feature distribution in Transferred Terrains.

5.6 Perceived Terrain Quality Metric (PTQM)

The results described above suggest that an automatic visual quality metric for terrains may be devised using geomorphons. To this end, we exploited a linear regression model, which predicts an overall perceptual terrain quality based on geomorphon features.

Correlation From the Pearson Correlation Coefficients table for our data in Table 5.4, we can see that there is a strong correlation between each of the geomorphons (our predictor variables) at various levels (Positive and Negative Correlation) on the Perception Score. The order of the influence on the perception score is given by: Valley (0.66), Ridge (0.64), Summit (0.44), Depression (0.42), Spur (0.33), Hollow (0.22), Flat (-0.10), Foot (-0.15), Shoulder (-0.17) and Slope (-0.65).
Perceived Terrain Quality Metric (PTQM) We performed a multiple linear regression (MLR) model on our dataset with the hypothesis, H_0 as follows: *"There is no linear relationship between the 10 geomorphon landform categories and the perception scores for our terrain data groups."*. The regression model is given by:

$$\begin{aligned} Y_{score} = {} & \beta_{int} + \beta_{depression}X_{depression} + \beta_{summit}X_{summit} + \\ & \beta_{flat}X_{flat} + \beta_{valley}X_{valley} + \beta_{ridge}X_{ridge} + \\ & \beta_{hollow}X_{hollow} + \beta_{spur}X_{spur} + \beta_{shoulder}X_{shoulder} + \\ & \beta_{slope}X_{slope} + \beta_{footslope}X_{footslope} + \sigma(Y), \end{aligned} \tag{5.1}$$

where $SD(Y) = \sigma$ that is independent from predictors. The regression gave us the following statistics: $DFn = 10, DFd = 588, F = 153.5276, p - value < 0.01$, and with $\alpha = 0.01$. Therefore, we rejected the null hypothesis. All coefficients are statistically significant with a $p - value$ less than 0.01. We get the coefficients: $\beta_{int} = -38.02$, $\beta_{depression} = 3.55$, $\beta_{summit} = 1.75$, $\beta_{flat} = 25.12$, $\beta_{valley} = 9.61$,

$\beta_{ridge} = 7.59$, $\beta_{hollow} = 6.71$, $\beta_{spur} = 9.02$, $\beta_{shoulder} = 7.31$, $\beta_{slope} = 28.95$, and $\beta_{footslope} = 7.63$.

The resulting R-Squared value for our regression model is 0.72 signifiying that the 72% of variation in the visual plausibility of terrains the perception score can be explained by the full model with all of our predictor variables 10 geomorphon distribution values with a standard error of 0.13. All of the landform factors are significant predictors of the perception score.

Based on our linear regression model between 10 geomorphons categories and the perception score we introduce a new metric to predict visual plausibility score for terrains, Perceived Terrain Quality Metrics (PTQM). The scale for the metric is $\langle 0.0, 1.0\rangle$ and higher is perceived as more visually plausible.

By substituting the values of geomorphons into the linear regression model in Eqn (5.2) we receive the PTQM:

$$\begin{aligned} PTQM \quad = \quad & -38.02 + 3.55G_{depression} + 1.75G_{summit} + \\ & 25.12G_{flat} + 9.61G_{valley} + 7.59G_{ridge} + \\ & 6.71G_{hollow} + 9.02G_{spur} + 7.31G_{shoulder} + \\ & 28.95G_{slope} + 7.63G_{footslope}. \end{aligned} \tag{5.2}$$

Table 5.9 and Figure 5.16 shows the comparison of PTQM with the calculated perception score averages for each category. The results from the same are presented below. The average of all real terrains PTQM=0.68 and synthetic PTQM=0.32. The SR is an outlier in synthetic terrains with PQTM=0.02 and if it is excluded, the average PTQM=0.38 for the synthetic terrain group 5.15.

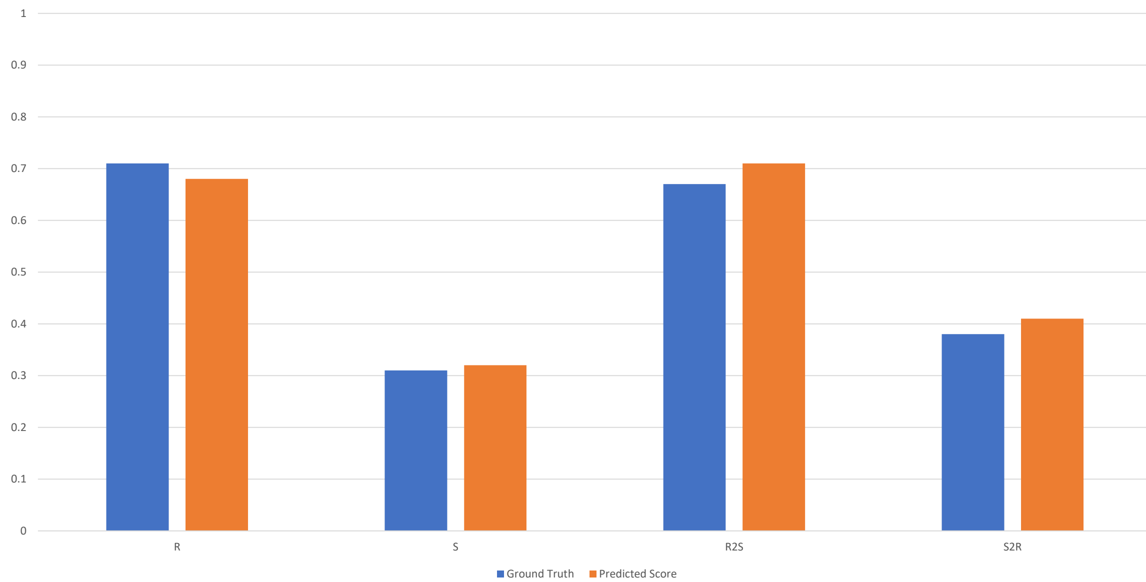


Figure 5.15. A comparison of the average value of measured perception scores (Ground Truth) for the overall terrain categories vs. the PTQM (Predicted Scores).

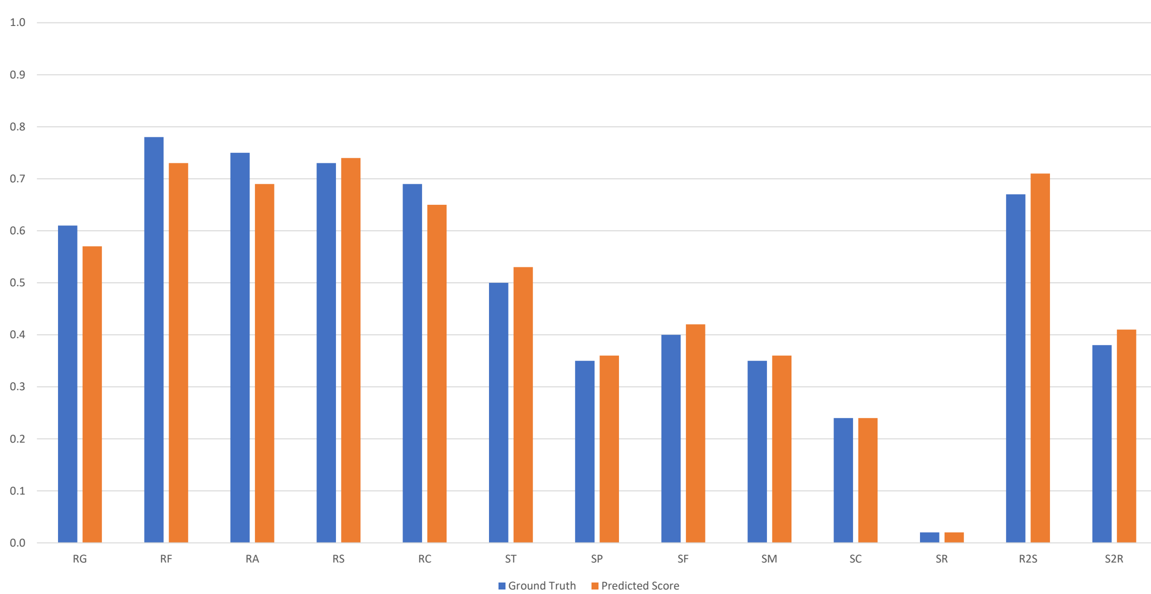


Figure 5.16. A comparison of the average value of measured perception scores (Ground Truth) for every terrain subcategories vs. the PTQM (Predicted Scores).

Table 5.7.

The correlations among ten geomorphons (Depression, Summit, Flat, Valley, Ridge, Hollow, Spur, Shoulder, Slope, and Footslope) and the perception scores.

CORR.	DEPR.	SUMM.	FLAT	VALL.	RIDG.	HOLL.	SPUR	SHOU.	SLOP.	FOOT.	SCORE
DEPR.	1.00	•	•	•	•	•	•	•	•	•	•
SUMM.	0.99	1.00	•	•	•	•	•	•	•	•	•
FLAT	-0.41	-0.41	1.00	•	•	•	•	•	•	•	•
VALL.	0.85	0.87	-0.42	1.00	•	•	•	•	•	•	•
RIDG.	0.86	0.87	-0.42	1.00	1.00	•	•	•	•	•	•
HOLL.	0.41	0.42	-0.77	0.49	0.49	1.00	•	•	•	•	•
SPUR	0.45	0.46	-0.76	0.56	0.57	0.99	1.00	•	•	•	•
SHOU.	-0.50	-0.51	0.71	-0.53	-0.54	-0.94	-0.93	1.00	•	•	•
SLOP.	-0.51	-0.53	-0.32	-0.67	-0.66	0.18	0.08	-0.18	1.00	•	•
FOOT	-0.50	-0.50	0.72	-0.51	-0.52	-0.95	-0.93	1.00	-0.20	1.00	•
SCORE	0.42	0.44	-0.10	0.66	0.64	0.22	0.33	-0.17	-0.65	-0.15	1.00

Table 5.8.

A comparison of perception scores generated based on our introduced metric and our previously normalized score from the study for the overall terrain groups.

Type	**Measured Perception Score**	**PTQM**
Real (R)	0.71	0.68
Synthetic (S)	0.31	0.32
Real Features to Synthetic Terrains (R2S)	0.67	0.71
Real Features to Synthetic Terrains (R2S)	0.38	0.41

Table 5.9.

A comparison of perception scores generated based on our introduced metric and our previously normalized score from the study for all the individual terrain subcategories.

Type	**Category**	**Measured Perception Score**	**PTQM**
Real (R)	RG	0.61	0.57
	RF	0.78	0.73
	RA	0.75	0.69
	RS	0.73	0.74
	RC	0.69	0.65
Synthetic (S)	ST	0.50	0.53
	SP	0.35	0.36
	SF	0.40	0.42
	SM	0.35	0.36
	SC	0.24	0.24
	SR	0.02	0.02
Transfers (T)	R2S	0.67	0.71
	S2R	0.38	0.41

5.7 Evaluation

We evaluate PTQM model by splitting the data five times randomly into 80:20%, recalculating the metrics on the 80% and validating on the 20%. The average regression equation from the evaluation is given by -38.44 + 3.61*Depression + 1.77*Summit + 25.40*Flat + 9.71*Valley + 7.65*Ridge + 6.77*Hollow + 9.14*Spur + 7.40*Shoulder + 29.26*Slope + 7.69*Footslope similar to the introduced PTQM model. The amount of explained variation (72%) and standard error (0.13) remained consistent as the regression model in the manuscript with 95% confidence interval. Table 5.10 shows the statistical and regression details of our validation technique. Figure 5.17, Figure 5.18, Figure 5.19, Figure 5.20, and Figure 5.21 shows the comparison of calculated scores from the evaluation and the measured score averages for each category. The match and similarity shows the robustness and validity of our method.

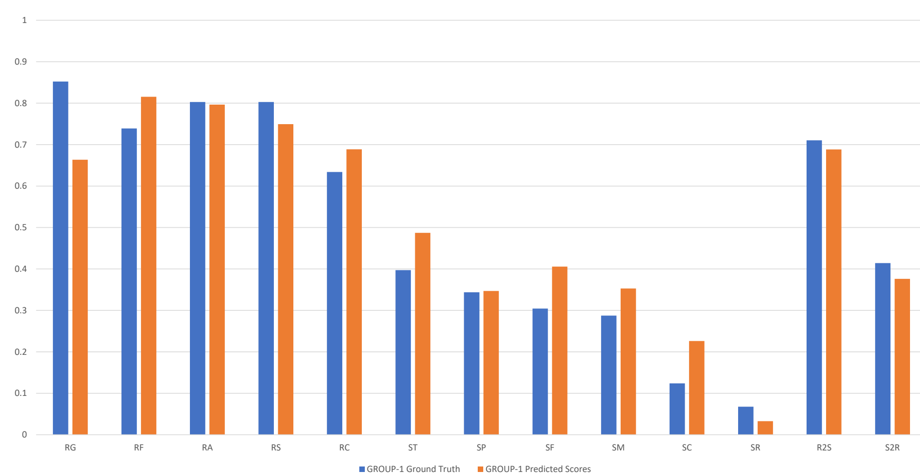


Figure 5.17. A comparison of measured scores (Ground Truth) and calculated scores (Predicted Scores) from Group-1 of Five random 80%:20% split regression evaluations.

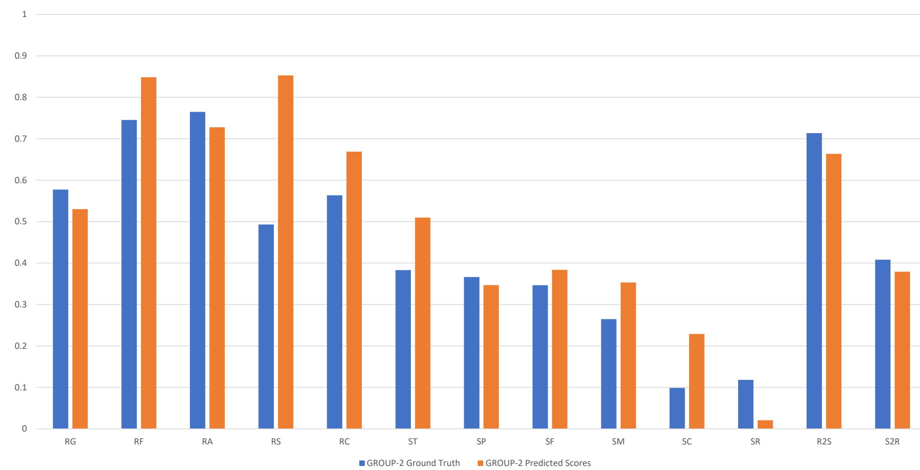


Figure 5.18. A comparison of measured scores (Ground Truth) and calculated scores (Predicted Scores) from Group-2 of Five random 80%:20% split regression evaluations.

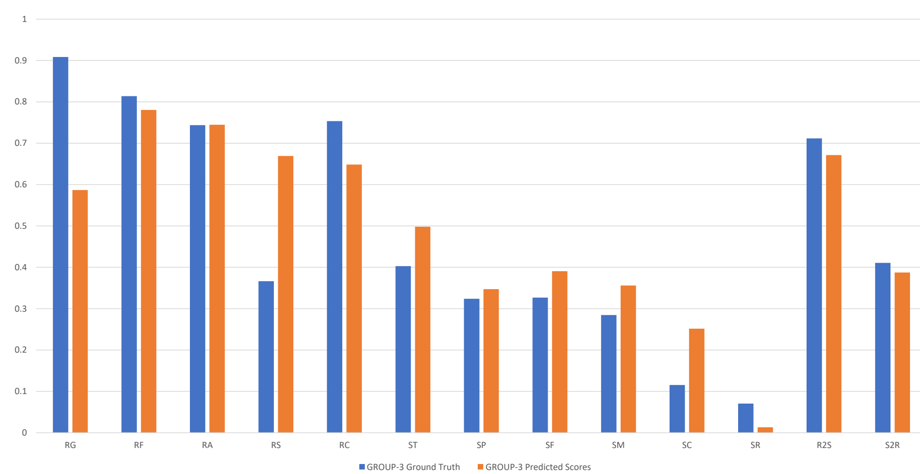


Figure 5.19. A comparison of measured scores (Ground Truth) and calculated scores (Predicted Scores) from Group-1 of Five random 80%:20% split regression evaluations.

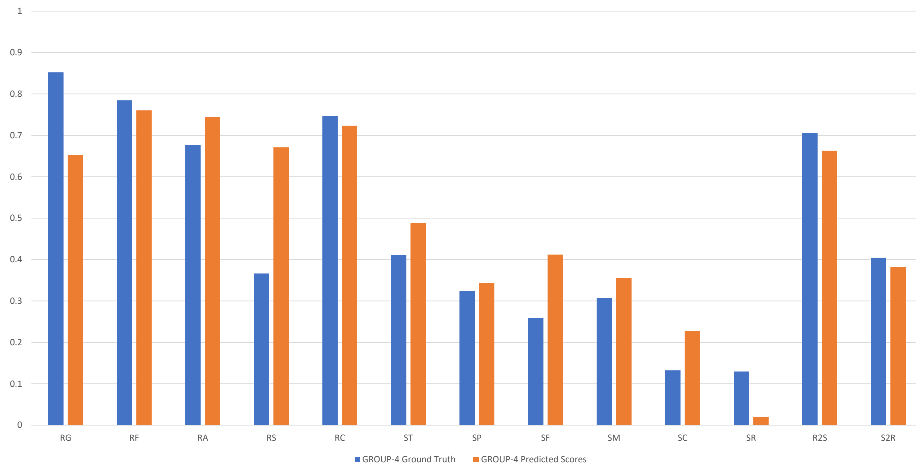


Figure 5.20. A comparison of measured scores (Ground Truth) and calculated scores (Predicted Scores) from Group-4 of Five random 80%:20% split regression evaluations.

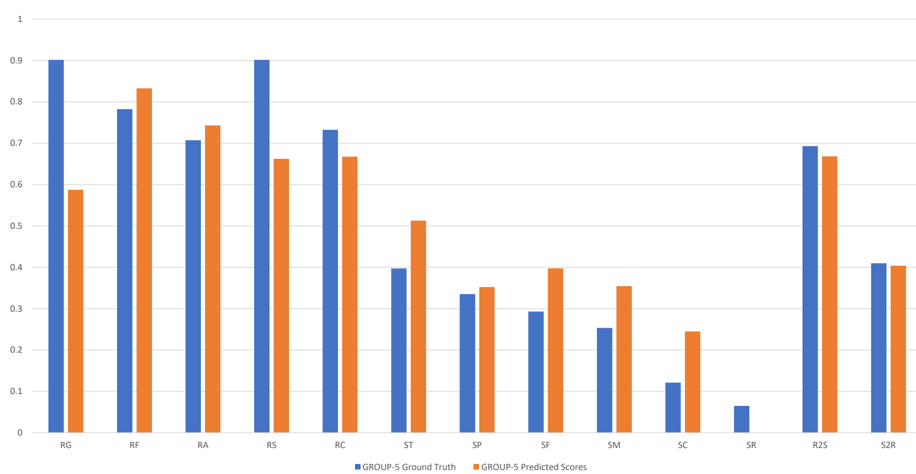


Figure 5.21. A comparison of measured scores (Ground Truth) and calculated scores (Predicted Scores) from Group-5 of Five random 80%:20% split regression evaluations.

Table 5.10.

Statistics and Regression coefficients for our 5 randomly split groups for evaluating PTQM.

Coefficients	**Group-1**	**Group-2**	**Group-3**	**Group-4**	**Group-5**
DF (Numerator)	10	10	10	10	10
DF (Denominator)	469	469	469	469	469
F	118.8014878	136.190991	121.7065162	122.6124249	128.5753699
p-value	<0.01	<0.01	<0.01	<0.01	<0.01
alpha	0.05	0.05	0.05	0.05	0.05
Adjusted R-Square	0.71	0.74	0.72	0.72	0.73
Standard Error	0.13	0.13	0.13	0.13	0.13
Intercept	-41.47	-41.54	-33.86	-38.05	-37.28
Depression	3.95	3.95	3.04	3.55	3.54
Summit	1.91	1.90	1.63	1.73	1.70
Flat	27.38	27.40	22.43	25.16	24.65
Valley	10.63	10.09	8.61	10.02	9.21
Ridge	8.08	8.62	6.75	7.18	7.63
Hollow	7.68	7.18	5.77	6.89	6.33
Spur	9.45	9.95	8.29	8.91	9.10
Shoulder	7.80	7.98	6.58	7.24	7.40
Slope	31.59	31.65	25.77	28.93	28.38
Footslope	8.46	8.31	6.75	7.71	7.24

CHAPTER 6. CONCLUSION

This paper presented a first step in the direction of evaluating the perceptual quality of procedural models of terrains. We have conducted two large scale perceptual studies on the Mechanical Turk surveying platform that allowed us to rank both synthetic and real terrains. The experiments show that synthetic terrains lack in visual quality. Our results indicate that synthetic terrains are perceived worse than real terrains with strong statistical significance.

We have performed a quantitative analysis on the terrains used in the study by using Geomorphons, a machine vision based landform classification system for DEMs that indicate geomorphological features such as Valley (0.66), Ridge (0.64), Summit (0.44), Depression (0.42), Spur(0.33), and Hollow (0.22) have significant perceptual importance in that order of influence on perceptual significance.

Then, by using multiple linear regression, we show that the presented geomorphon features are strongly correlated with perceived visual quality and inturn the perception scores. We used deep neural networks, specifically, a Bicycle GAN to transfer the features and the second perceptual study confirmed this observation. Eventually, we have designed a novel perceptual metrics based on geomorphons that allows to assign a number of estimated visual quality of the generated terrain.

Based on the conducted experiments and statistics, we can answer our research questions as follows:

- Procedural terrain models are not perceived equally as good as the real terrains.
- The geomorphon features such as Valley, Ridge, Summit, Depression, Spur and Hollow have the strongest influence on perceived quality of the terrain models.

- It is possible to transfer features from procedural terrains to real terrains and vice versa to study their underlying perceptual phenomenon using deep neural networks especially Bicycle GANs but we do not have precise control over the features that will be transferred.

6.1 Contributions

In this body of research, we claim the following contributions:

1. We introduce Perceived Terrain Quality Metrics, a novel procedural terrain evaluation metric that assigns a normalized value of perceived perception to a terrain represented as a digital elevation model.

2. We have conducted large scale user studies that confirm and validate that procedural terrains are not visually plausible as real terrains and to measure the visual plausibility of real and synthetic terrain models.

3. We have determined the geomorphological features that have the strongest effect on visual plausibility of terrains.

4. We provide a publicly available data-set of real and procedural terrains with assigned perceptual evaluation scores and calculated geomorphons along with their individual statistics.

6.2 Limitations

Our study has several **limitations** that are addressed as follows:

1. Geomorphons are localized to small areas of the terrain and they do not reflect the distributions of the large features such as rivers, large valleys, etc. It is possible that two terrains with the same feature vector may be perceived as different because of the variety of distributions and their presence in conjunction with some other features that has not been studied yet.

2. Our research design made several assumptions on size of the terrains. The changing of the scale on these terrains may have different effects on our results because geomorphological features of different scales would be captured by magnification or minification of the terrain scale.

3. Another important limitation is the assumption about the terrain classification while we motivated our classification into terrains with different geomorphological patterns, it is well known that probably every terrain on Earth has been exposed to eons of various morphing phenomena through erosion and weathering. Therefore, it is not entirely clear what exactly caused those patterns as the result of each geological feature is concatenation of multiple geomorphological processes acting in parallel or in a sequence.

4. We also assumed a fixed position of the camera, consistent texturing, and illumination. While these aspects were carefully selected and made constant, it would be interesting to see the effect of each of them on the results such as using a rendering style or grayscale textures.

5. The deep learning based feature transfer with GAN provides limited control on the content to be or not to be transferred. With the metric we provided, the transferred results can be further improved in perception with a better control schema of the generative network.

6. Lastly, We also did not study the spatial correlation between geomorphon features.

6.3 Recommendations

The order of the influence on the perception score is given by: Valley (0.66), Ridge (0.64), Summit (0.44), Depression (0.42), Spur (0.33), Hollow (0.22), Flat (-0.10), Foot (-0.15), Shoulder (-0.17) and Slope (-0.65). Based on our results from E-2 (Section 4.2.3), each of our procedurally generated terrain model categories does not have enough variety of geomorphogical features 5.13 as the real terrain models 5.12 which is highly evident from our deep learning based feature transfer process. Therefore, when generating a procedural terrain in CG, using a multitude of erosion methods will result in generating more geomorphological variety in these terrains which will inturn increased the perceived quality.

6.4 Future Work

There are many possible avenues for **future work**:

1. Perceptual studies have the potential to answer longstanding questions of visual quality of procedural models. Our work is based on the underlying concept of geomorphons that may be difficult to generalize to different domains. It would be interesting to develop similar metrics for vegetation, urban models, etc. Generalizing this work to other simulations, such as fluids, would be also an interesting future work.

2. Augmenting the current research methodology with qualitative research and design workshops to gather other levels of information may provide more meaningful insights on the aspects of the attributes we are looking for in procedural terrains that are typically non-quantifiable by geomorphons.

3. Consideration of other non-intrinsic properties that may contribute and affect the perception of a terrain such as the Environmental atttibutes and context based conditions.

6.5 Summary

In this chapter of the dissertation, we summarized our methods, presented our final results and the novel perceptual evalution metric we have formulated for terrains namely, Perceptual Terrain Quality Metric in computer graphics. In addition, we also addressed the limitations of our proposed methods in augmentation with a list of recommendations for future terrain modeling including some of the possible venues for future work and extending the proposed method to other categories of procedural models in CG.

LIST OF REFERENCES

Almahairi, A., Rajeswar, S., Sordoni, A., Bachman, P., & Courville, A. (2018). Augmented cyclegan: Learning many-to-many mappings from unpaired data. *arXiv preprint arXiv:1802.10151*.

Anh, N. H., Sourin, A., & Aswani, P. (2007). Physically based hydraulic erosion simulation on graphics processing unit. In *Proceedings of the graphite* (pp. 257–264). New York, NY, USA: ACM.

Aydın, T. O., Čadík, M., Myszkowski, K., & Seidel, H.-P. (2010). Video quality assessment for computer graphics applications. In *Acm transactions on graphics (proc. of siggraph asia)* (pp. 1–10). Seoul, Korea: ACM.

Aydın, T. O., Smolic, A., & Gross, M. (2015, Jan). Automated aesthetic analysis of photographic images. *IEEE Transactions on Visualization and Computer Graphics*, *21*(1), 31-42. doi: 10.1109/TVCG.2014.2325047

Bartz, D., Cunningham, D. W., Fischer, J., & Wallraven, C. (2008). The role of perception for computer graphics. In *Eurographics (stars)* (pp. 59–80).

Benes, B., & Forsbach, R. (2002). Visual simulation of hydraulic erosion. *Journal of WSCG*, *10*(1), 79–86.

Benes, B., Těšínský, V., Hornyš, J., & Bhatia, S. K. (2006). Hydraulic erosion. *Computer Animation and Virtual Worlds*, *17*(2), 99–108. doi: http://dx.doi.org/10.1002/cav.v17:2

Bojrab, M., Abdul-Massih, M., & Benes, B. (2013, March). Perceptual importance of lighting phenomena in rendering of animated water. *ACM Trans. Appl. Percept.*, *10*(1), 2:1–2:18. Retrieved from `http://doi.acm.org/10.1145/2422105.2422107` doi: 10.1145/2422105.2422107

Bullard, J. E., & Livingstone, I. (2002). Interactions between aeolian and fluvial systems in dryland environments. *Area*, *34*(1), 8–16.

Chen, T. Q., & Schmidt, M. (2016). Fast patch-based style transfer of arbitrary style. *arXiv preprint arXiv:1612.04337*.

Chen, X., Saparov, A., Pang, B., & Funkhouser, T. (2012, July). Schelling points on 3d surface meshes. *ACM Trans. Graph.*, *31*(4), 29:1–29:12. Retrieved from `http://doi.acm.org/10.1145/2185520.2185525` doi: 10.1145/2185520.2185525

Cheng, I., & Boulanger, P. (2005). A 3d perceptual metric using just-noticeable-difference. In *Eurographics (short presentations)* (pp. 97–100).

Cheng, I., Firouzmanesh, A., & Basu, A. (2011). Perceptual factors in graphics: from jnd to pam. In *Computer vision, pattern recognition, image processing and graphics (ncvpripg), 2011 third national conference on* (pp. 6–10).

Cherian, A., & Sullivan, A. (2019). Sem-gan: Semantically-consistent image-to-image translation. In *2019 ieee winter conference on applications of computer vision (wacv)* (pp. 1797–1806).

Cleju, I., & Saupe, D. (2006). Evaluation of supra-threshold perceptual metrics for 3d models. In *Proceedings of the 3rd symposium on applied perception in graphics and visualization* (pp. 41–44).

Cordonnier, G., Braun, J., Cani, M.-P., Benes, B., Galin, E., Peytavie, A., & Guérin, E. (2016). Large scale terrain generation from tectonic uplift and fluvial erosion. In *Computer graphics forum* (Vol. 35, pp. 165–175).

Cordonnier, G., Cani, M.-P., Benes, B., Braun, J., & Galin, E. (2018, May). Sculpting mountains: Interactive terrain modeling based on subsurface geology. *IEEE Transactions on Visualization and Computer Graphics*, *24*(5), 1756-1769. doi: 10.1109/TVCG.2017.2689022

Cordonnier, G., Galin, E., Gain, J., Benes, B., Guérin, E., Peytavie, A., & Cani, M.-P. (2017). Authoring landscapes by combining ecosystem and terrain erosion simulation. *ACM Transactions on Graphics*, *36*(4).

Corsini, M., Larabi, M.-C., Lavoué, G., Petřík, O., Váša, L., & Wang, K. (2013). Perceptual metrics for static and dynamic triangle meshes. In *Computer graphics forum* (Vol. 32, pp. 101–125).

Daniel, T. C. (2001). Whither scenic beauty? visual landscape quality assessment in the 21st century. *Landscape and Urban Planning*, *54*(1), 267 - 281. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0169204601001414` (Our Visual Landscape: analysis, modeling, visualization and protection) doi: https://doi.org/10.1016/S0169-2046(01)00141-4

Draguţ, L., & Blaschke, T. (2006, 11). Automated classification of landform elements using object-based image analysis. *Geomorphology*, *81*, 330-344. doi: 10.1016/j.geomorph.2006.04.013

Drascic, D., & Milgram, P. (1996). Perceptual issues in augmented reality. In *Proceedings-spie the international society for optical engineering* (pp. 123–134).

Dünser, A., Grasset, R., & Billinghurst, M. (2008). *A survey of evaluation techniques used in augmented reality studies*. Human Interface Technology Laboratory New Zealand.

Elloumi, N., Kacem, H. L. H., Dey, N., Ashour, A. S., & Bouhlel, M. S. (2017). Perceptual metrics quality: Comparative study for 3d static meshes. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, *8*(1), 63–80.

E-onSoftware. (2019). *Home.* Retrieved from `https://info.e-onsoftware.com/home`

Farr, T. G., & Kobrick, M. (2000). Shuttle radar topography mission produces a wealth of data. *Eos, Transactions American Geophysical Union*, *81*(48), 583–585.

Ferwerda, J. A., Pattanaik, S. N., Shirley, P., & Greenberg, D. P. (1996). A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd annual conference on computer graphics and interactive techniques* (pp. 249–258). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/237170.237262` doi: 10.1145/237170.237262

Ferwerda, J. A., Shirley, P., Pattanaik, S. N., & Greenberg, D. P. (1997). A model of visual masking for computer graphics. In *Proceedings of the 24th annual conference on computer graphics and interactive techniques* (pp. 143–152). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co. Retrieved from `https://doi.org/10.1145/258734.258818` doi: 10.1145/258734.258818

Fournier, A., Fussell, D., & Carpenter, L. (1982). Computer rendering of stochastic models. *Computer Graphics*, *25*(6), 371–384.

Frintrop, S., Werner, T., & Garca, G. M. (2015, June). Traditional saliency reloaded: A good old model in new shape. In *2015 ieee conference on computer vision and pattern recognition (cvpr)* (p. 82-90). doi: 10.1109/CVPR.2015.7298603

Gain, J., Merry, B., & Marais, P. (2015). Parallel, realistic and controllable terrain synthesis. *Computer Graphics Forum*, *34*(2), 105–116.

Gain, J. E., Marais, P., & Strasser, W. (2009). Terrain sketching. In *Proceedings of the symposium on interactive 3d graphics and games* (pp. 31–38). Boston, USA: ACM.

Galin, E., Guérin, E., Peytavie, A., Cordonnier, G., Cani, M.-P., Benes, B., & Gain, J. (2019). A Review of Digital Terrain Modeling. *Computer Graphics Forum*, *38*(2).

Gan, Y., Cai, X., Liu, J., & Wang, S. (2015). A texture retrieval scheme based on perceptual features. In *Signal and information processing association annual summit and conference (apsipa), 2015 asia-pacific* (pp. 897–900).

Gan, Y., Chi, H., Gao, Y., Liu, J., Zhong, G., & Dong, J. (2017). Perception driven texture generation. *arXiv preprint arXiv:1703.09784*.

Gao, X., Brooks, S., & Arnold, D. V. (2017). A feature-based quality metric for tone mapped images. *ACM Transactions on Applied Perception (TAP)*, *14*(4), 26.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.

Gibson, J. J. (1950). The perception of the visual world.

Gooch, A., Gooch, B., Shirley, P., & Cohen, E. (1998). A non-photorealistic lighting model for automatic technical illustration. In *Proceedings of the 25th annual conference on computer graphics and interactive techniques* (pp. 447–452).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).

Goral, C. M., Torrance, K. E., Greenberg, D. P., & Battaile, B. (1984). Modeling the interaction of light between diffuse surfaces. In *Acm siggraph computer graphics* (Vol. 18, pp. 213–222).

Guérin, E., Digne, J., Galin, E., & Peytavie, A. (2016). Sparse representation of terrains for procedural modeling. *Computer Graphics Forum (Proceedings of Eurographics)*, *35*(2), 177–187.

Guérin, E., Digne, J., Galin, E., Peytavie, A., Wolf, C., Benes, B., & Martinez, B. (2017). Interactive example-based terrain authoring with conditional generative adversarial networks. *ACM Transactions on Graphics (Proceedings of Siggraph Asia 2017)*, *36*(6).

Guo, J., Vidal, V., Baskurt, A., & Lavoué, G. (2015). Evaluating the local visibility of geometric artifacts. In *Proceedings of the acm siggraph symposium on applied perception* (pp. 91–98). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2804408.2804418` doi: 10.1145/2804408.2804418

Guo, J., Vidal, V., Cheng, I., Basu, A., Baskurt, A., & Lavoue, G. (2016, October). Subjective and objective visual quality assessment of textured 3d meshes. *ACM Trans. Appl. Percept.*, *14*(2), 11:1–11:20. Retrieved from `http://doi.acm.org/10.1145/2996296` doi: 10.1145/2996296

Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., & Gool, L. V. (2013, Dec). The interestingness of images. In *2013 ieee international conference on computer vision* (p. 1633-1640). doi: 10.1109/ICCV.2013.205

Hagerhall, C. M., Purcell, T., & Taylor, R. (2004). Fractal dimension of landscape silhouette outlines as a predictor of landscape preference. *Journal of Environmental Psychology*, *24*(2), 247 - 255. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0272494404000076` doi: https://doi.org/10.1016/j.jenvp.2003.12.004

Hamill, J., McDonnell, R., Dobbyn, S., & O'Sullivan, C. (2005). Perceptual evaluation of impostor representations for virtual humans and buildings. In *Computer graphics forum* (Vol. 24, pp. 623–633).

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017, Oct). Mask r-cnn. In *The ieee international conference on computer vision (iccv).*

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Herzog, R., Čadík, M., Aydın, T. O., Kim, K. I., Myszkowski, K., & Seidel, H.-P. (2012). NoRM: no-reference image quality metric for realistic image synthesis. *Computer Graphics Forum*, *31*(2), 545–554. doi: 10.1111/j.1467-8659.2012.03055.x

Hnaidi, H., Guérin, É., Akkouche, S., Peytavie, A., & Galin, É. (2010). Feature based terrain generation using diffusion equation. *Computer Graphics Forum*, *29*(7), 2179–2186.

Hodgins, J., Jörg, S., O'Sullivan, C., Park, S. I., & Mahler, M. (2010). The saliency of anomalies in animated human characters. *ACM Transactions on Applied Perception (TAP)*, *7*(4), 22.

Hoffman, B., & Nadelson, L. (2010). Motivational engagement and video gaming: A mixed methods study. *Educational Technology Research and Development*, *58*(3), 245–270.

Huggett, R. (2016). *Fundamentals of geomorphology*. Routledge.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2016). Image-to-image translation with conditional adversarial networks. *arxiv*.

Jasiewicz, J., & Stepinski, T. F. (2013). Geomorphons - a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, *182*, 147 - 156. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0169555X12005028` doi: https://doi.org/10.1016/j.geomorph.2012.11.005

Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2017). Neural style transfer: A review. *arXiv preprint arXiv:1705.04058*.

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision* (pp. 694–711).

Jones, J. A., Swan II, J. E., Singh, G., Kolstad, E., & Ellis, S. R. (2008). The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th symposium on applied perception in graphics and visualization* (pp. 9–14).

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. In *Computer vision, 2009 ieee 12th international conference on* (pp. 2106–2113).

Kamada, T., & Kawai, S. (1988). A simple method for computing general position in displaying three-dimensional objects. *Computer Vision, Graphics, and Image Processing*, *41*(1), 43–56.

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.

Kelley, A. D., Malin, M. C., & Nielson, G. M. (1988). Terrain simulation using a model of stream erosion. *Computer Graphics*, *22*(4), 263–268.

Kim, T., Cha, M., Kim, H., Lee, J. K., & Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. In *Proceedings of the 34th international conference on machine learning - volume 70* (pp. 1857–1865). JMLR.org. Retrieved from `http://dl.acm.org/citation.cfm?id=3305381.3305573`

King, N., et al. (2004). Using templates in the thematic analysis of text. *Essential guide to qualitative methods in organizational research*, *2*, 256–70.

Krištof, P., Benes, B., Křivánek, J., & Šťava, O. (2009). Hydraulic erosion using smoothed particle hydrodynamics. *Computer Graphics Forum*, *28*(2). Retrieved from `http://www2.tech.purdue.edu/cgt/Facstaff/bbenes/private/papers/EG09SPH.zip`

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 25* (pp. 1097–1105). Curran Associates, Inc. Retrieved from `http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf`

Kruijff, E., Swan, J. E., & Feiner, S. (2010). Perceptual issues in augmented reality revisited. In *Mixed and augmented reality (ismar), 2010 9th ieee international symposium on* (pp. 3–12).

Landy, M. S., & Graham, N. (2004). 73 visual perception of texture. *The visual neurosciences*, *1*, 1106.

Lavoué, G. (2011). A multiscale metric for 3d mesh visual quality assessment. In *Computer graphics forum* (Vol. 30, pp. 1427–1437).

Lavoué, G., Larabi, M., & Vása, L. (2016). On the efficiency of image metrics for evaluating the visual quality of 3d models. *IEEE Trans. Vis. Comput. Graph.*, *22*(8), 1987–1999.

LeCun, Y., Bengio, Y., & Hinton, G. (2015a). Deep learning. *nature*, *521*(7553), 436.

LeCun, Y., Bengio, Y., & Hinton, G. (2015b). Deep learning. *nature*, *521*(7553), 436.

Lee, C. H., Varshney, A., & Jacobs, D. W. (2005). Mesh saliency. In *Acm transactions on graphics (tog)* (Vol. 24, pp. 659–666).

Li, C., & Wand, M. (2016). Combining markov random fields and convolutional neural networks for image synthesis. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2479–2486).

Li, Q., Wang, Z., & Yang, B. (2008). Multi-resolution representation of digital terrain models with terrain features preservation. *Science in China Series E: Technological Sciences*, *51*, 145–154.

Liao, W.-H. (2010). Region description using extended local ternary patterns. In *2010 20th international conference on pattern recognition* (pp. 1003–1006).

Lin, W., & Kuo, C.-C. J. (2011). Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, *22*(4), 297–312.

Lindsay, J. B. (2016). Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. *Hydrological Processes*, *30*(6), 846–857.

Liu, J., Dong, J., Cai, X., Qi, L., & Chantler, M. (2015a). Visual perception of procedural textures: Identifying perceptual dimensions and predicting generation models. *PloS one*, *10*(6), e0130335.

Liu, J., Dong, J., Cai, X., Qi, L., & Chantler, M. (2015b, 06). Visual perception of procedural textures: Identifying perceptual dimensions and predicting generation models. *PLOS ONE*, *10*(6), 1-22. Retrieved from `https://doi.org/10.1371/journal.pone.0130335` doi: 10.1371/journal.pone.0130335

Liu, J., Dong, J., Qi, L., & Chantler, M. (2013). Identifying perceptual features of procedural textures. *Perception*, *42*, 221–221.

Loomis, J. M., & Knapp, J. M. (2003). Visual perception of egocentric distance in real and virtual environments. *Virtual and adaptive environments*, *11*, 21–46.

Luan, F., Paris, S., Shechtman, E., & Bala, K. (2017, July). Deep photo style transfer. In *The ieee conference on computer vision and pattern recognition (cvpr).*

Mandelbrot, B. B. (1988). The science of fractal images. In (1st ed., pp. 243–260). Springer.

Mantiuk, R., Kim, K. J., Rempel, A. G., & Heidrich, W. (2011, July). Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, *30*(4), 40:1–40:14. Retrieved from `http://doi.acm.org/10.1145/2010324.1964935` doi: 10.1145/2010324.1964935

Mantiuk, R., Myszkowski, K., & Seidel, H.-P. (2006, July). A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept.*, *3*(3), 286–308. Retrieved from `http://doi.acm.org/10.1145/1166087.1166095` doi: 10.1145/1166087.1166095

McNamara, A., Mania, K., & Gutierrez, D. (2011). Perception in graphics, visualization, virtual environments and animation. In *Siggraph asia 2011 courses* (p. 17).

Miller, G. (1986). The definition and rendering of terrain maps. *Computer Graphics*, *20*(4), 39–48.

Miller, K. (2005). *Communication theories: Perspectives, processes, and contexts (ediţia a 2-a).* McGraw-Hill, Londra.

Musgrave, F. K., Kolb, C. E., & Mace, R. S. (1989). The synthesis and rendering of eroded fractal terrains. *Computer Graphics*, *23*(3), 41–50.

Musialski, P., Wonka, P., Aliaga, D. G., Wimmer, M., Gool, L. v., & Purgathofer, W. (2013). A survey of urban reconstruction. In *Computer graphics forum* (Vol. 32, pp. 146–177).

Myszkowski, K. (2002). Perception-based global illumination, rendering, and animation techniques. In *Proceedings of the 18th spring conference on computer graphics* (pp. 13–24).

Nader, G., Wang, K., Htroy-Wheeler, F., & Dupont, F. (2016). Visual contrast sensitivity and discrimination for 3d meshes and their applications. *Computer Graphics Forum*, *35*(7), 497-506. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13046` doi: 10.1111/cgf.13046

Neidhold, B., Wacker, M., & Deussen, O. (2005). Interactive physically based fluid and erosion simulation. In *Proceedings of eurographics conference on natural phenomena* (pp. 25–33). Retrieved from `http://dx.doi.org/10.2312/NPH/NPH05/025-032` doi: 10.2312/NPH/NPH05/025-032

Neteler, M., & Mitasová, H. (2013). *Open source gis: a grass gis approach* (Vol. 689). Springer Science & Business Media.

Neumann, L., Čadík, M., & Nemcsics, A. (2007). An efficient perception-based adaptive color to gray transformation. In *Proceedings of computational aesthetics 2007* (pp. 73– 80). Banff, Canada: Eurographics Association.

Neumann, L., Sbert, M., Gooch, B., Purgathofer, W., et al. (2005). Viewpoint quality: Measures and applications. In *Proceedings of the 1st eurographics workshop on computational aesthetics in graphics, visualization and imaging. aire-la-vile: The eurographics association press* (pp. 185–192).

Odena, A., Dumoulin, V., & Olah, C. (2016). Deconvolution and checkerboard artifacts. *Distill*. Retrieved from `http://distill.pub/2016/deconv-checkerboard` doi: 10.23915/distill.00003

Ondřej, J., Ennis, C., Merriman, N. A., & O'sullivan, C. (2016, July). Frankenfolk: Distinctiveness and attractiveness of voice and motion. *ACM Trans. Appl. Percept.*, *13*(4), 20:1–20:13. Retrieved from `http://doi.acm.org/10.1145/2948066` doi: 10.1145/2948066

O'Sullivan, C., Howlett, S., McDonnell, R., Morvan, Y., & O'Conor, K. (2004). Perceptually Adaptive Graphics. In *Eurographics 2004 - stars.* Eurographics Association. doi: 10.2312/egst.20041029

OSullivan, C., Howlett, S., Morvan, Y., McDonnell, R., & OConor, K. (2004). Perceptually adaptive graphics. *Eurographics state of the art reports*, *4*, 1–24.

Palmer, J. (2003). Research agenda for landscape perception. *Trends in Landscape Modelling, Proceedings at Anhalt University of Applied Sciences*.

Panetta, K. A., Wharton, E. J., & Agaian, S. S. (2008). Human visual system-based image enhancement and logarithmic contrast measure. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *38*(1), 174–188.

Peng, X. B., Berseth, G., & Van de Panne, M. (2016). Terrain-adaptive locomotion skills using deep reinforcement learning. *ACM Transactions on Graphics (TOG)*, *35*(4), 81.

Perez-Ortiz, M., & Mantiuk, R. K. (2017). A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint arXiv:1712.03686*.

Perlin, K. (1985). An image synthesizer. *SIGGRAPH Comput. Graph.*, *19*(3), 287–296.

Peytavie, A., Galin, E., Grosjean, J., & Mérillou, S. (2009). Arches: a framework for modeling complex terrains. In *Computer graphics forum* (Vol. 28, pp. 457–467).

PlanetsideSoftware. (2019). Retrieved from `https://planetside.co.uk/`

Reddy, M. (1997). *Perceptually modulated level of detail for virtual environments.* Unpublished doctoral dissertation, University of Edinburgh, UK.

Reddy, M. (2001, September). Perceptually optimized 3d graphics. *IEEE Comput. Graph. Appl.*, *21*(5), 68–75. Retrieved from `http://dx.doi.org/10.1109/38.946633` doi: 10.1109/38.946633

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*.

Reitsma, P. S. A., & Pollard, N. S. (2003, July). Perceptual metrics for character animation: Sensitivity to errors in ballistic motion. *ACM Trans. Graph.*, *22*(3), 537–542. Retrieved from `http://doi.acm.org/10.1145/882262.882304` doi: 10.1145/882262.882304

Riche, N., Duvinage, M., Mancas, M., Gosselin, B., & Dutoit, T. (2013, Dec). Saliency and human fixations: State-of-the-art and study of comparison metrics. In *2013 ieee international conference on computer vision* (p. 1153-1160). doi: 10.1109/ICCV.2013.147

Riehle, D. (2000). *Framework design: A role modeling approach.* Unpublished doctoral dissertation, ETH Zurich.

Rushmeier, H., Rogowitz, B., & Piatko, C. (2000, 06/2000). Perceptual issues in substituting texture for geometry. In *Spie: Human vision and electronic imaging* (Vol. 3959, p. 372-383). Retrieved from `http://dx.doi.org/10.1117/12.387174` doi: 10.1117/12.387174

Saghri, J. A., Cheatham, P. S., & Habibi, A. (1989). Image quality measure based on a human visual system model. *Optical Engineering*, *28*(7), 287813.

Schwarz, M., & Stamminger, M. (2009). On predicting visual popping in dynamic scenes. In *Proceedings of the 6th symposium on applied perception in graphics and visualization* (pp. 93–100). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/1620993.1621012` doi: 10.1145/1620993.1621012

Sedgwick, H. (1983). Environment-centered representation of spatial layout: Available visual information from texture and perspective. In *Human and machine vision* (pp. 425–458). Elsevier.

Shi, Y., Ondřej, J., Wang, H., & O'Sullivan, C. (2017, March). Shape up! perception based body shape variation for data-driven crowds. In *2017 ieee virtual humans and crowds for immersive environments (vhcie)* (p. 1-7). doi: 10.1109/VHCIE.2017.7935623

SideFX. (2019). Retrieved from `https://www.sidefx.com/`

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Smelik, R. M., De Kraker, K. J., Tutenel, T., Bidarra, R., & Groenewegen, S. A. (2009). A survey of procedural methods for terrain modelling. In *Proceedings of the casa workshop on 3d advanced media in gaming and simulation (3amigas)* (pp. 25–34).

Smelik, R. M., Tutenel, T., Bidarra, R., & Benes, B. (2014). A survey on procedural modelling for virtual worlds. In *Computer graphics forum* (Vol. 33, pp. 31–50).

Smith, K., Landes, P.-E., Thollot, J., & Myszkowski, K. (2008, April). Apparent Greyscale: A Simple and Fast Conversion to Perceptually Accurate Images and Video. *Computer Graphics Forum*, *27*(2), 193-200. Retrieved from `https://hal.inria.fr/inria-00255958` doi: 10.1111/j.1467-8659.2008.01116.x

SpeedTree. (2019). Retrieved from `https://store.speedtree.com/`

Spehar, B., Clifford, C. W. G., Newell, B. R., & Taylor, R. P. (2003). Universal aesthetic of fractals. *Computers & Graphics*, *27*(5), 813–820. Retrieved from `https://doi.org/10.1016/S0097-8493(03)00154-7` doi: 10.1016/S0097-8493(03)00154-7

Št'ava, O., Benes, B., Brisbin, M., & Křivánek, J. (2008). Interactive terrain modeling using hydraulic erosion. In *Proceedings of the 2008 acm siggraph/eurographics symposium on computer animation* (pp. 201–210).

St'ava, O., Benes, B., Brisbin, M., & Křivánek, J. (2008). Interactive terrain modeling using hydraulic erosion. In *Proceedings of the 2008 acm siggraph/eurographics symposium on computer animation* (pp. 201–210). Aire-la-Ville, Switzerland, Switzerland: Eurographics Association. Retrieved from `http://dl.acm.org/citation.cfm?id=1632592.1632622`

Taylor, R., Spehar, B., van Donkelaar, P., & Hgerhll, C. (2011, 06). Perceptual and physiological responses to jackson pollock's fractals. *Frontiers in human neuroscience*, *5*, 60. doi: 10.3389/fnhum.2011.00060

Thompson, W., Fleming, R., Creem-Regehr, S., & Stefanucci, J. K. (2011). *Visual perception from a computer graphics perspective*. CRC Press.

Toussaint, G. T. (2000). Complexity of computing nice viewpoints of objects in space. In *Proc spie int soc opt eng* (Vol. 4117, pp. 1–11).

Travers, R. M. (1984). Human information processing.

Tremblet, A. (2016). The mountain sublime of philip james de loutherbourg and joseph mallord william turner. *Journal of Alpine Research*, 104-2. doi: https://doi.org/10.4000/rga.3395

Tveit, M. S., Sang, A. O., & Hagerhall, C. M. (2012). Scenic beauty: Visual landscape assessment and human landscape perception. In L. Steg, A. E. van den Berg, & J. I. de Groot (Eds.), *Environmental psychology: an introduction* (p. 37-46). UK: BPS Blackwell.

Ulyanov, D., Lebedev, V., Vedaldi, A., & Lempitsky, V. S. (2016). Texture networks: Feed-forward synthesis of textures and stylized images. In *Icml* (Vol. 1, p. 4).

Um, K., Hu, X., & Thuerey, N. (2017). Perceptual evaluation of liquid simulation methods. *ACM Transactions on Graphics (TOG)*, *36*(4), 143.

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & health sciences*, *15*(3), 398–405.

Vanek, J., Benes, B., Herout, A., & Stava, O. (2011). Large-scale physics-based terrain editing using adaptive tiles on the gpu. *IEEE Computer Graphics and Applications*, *31*(6), 35–44.

Vanhoey, K., Sauvage, B., Kraemer, P., & Lavoué, G. (2017, October). Visual quality assessment of 3d models: On the influence of light-material interaction. *ACM Trans. Appl. Percept.*, *15*(1), 5:1–5:18. Retrieved from `http://doi.acm.org/10.1145/3129505` doi: 10.1145/3129505

Wang, H., Ondřej, J., & O'Sullivan, C. (2017, May). Trending paths: A new semantic-level metric for comparing simulated and real crowd data. *IEEE Transactions on Visualization and Computer Graphics*, *23*(5), 1454-1464. doi: 10.1109/TVCG.2016.2642963

Wang, H., Ondřej, J., & O'Sullivan, C. (2016). Path patterns: Analyzing and comparing real and simulated crowds. In *Proceedings of the 20th acm siggraph symposium on interactive 3d graphics and games* (pp. 49–57). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2856400.2856410` doi: 10.1145/2856400.2856410

Wang, Q., Zhang, L., Bertinetto, L., Hu, W., & Torr, P. H. (2019). Fast online object tracking and segmentation: A unifying approach. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004, April). Image quality assessment: From error visibility to structural similarity. *Trans. Img. Proc.*, *13*(4), 600–612. Retrieved from `http://dx.doi.org/10.1109/TIP.2003.819861` doi: 10.1109/TIP.2003.819861

Weier, M., Stengel, M., Roth, T., Didyk, P., Eisemann, E., Eisemann, M., Grogorick, S., ... others (2017a). Perception-driven accelerated rendering. In *Computer graphics forum* (Vol. 36, pp. 611–643).

Weier, M., Stengel, M., Roth, T., Didyk, P., Eisemann, E., Eisemann, M., Grogorick, S., … Slusallek, P. (2017b, May). Perception-driven accelerated rendering. *Comput. Graph. Forum*, *36*(2), 611–643. Retrieved from `https://doi.org/10.1111/cgf.13150` doi: 10.1111/cgf.13150

Winkler, S., & Mohandas, P. (2008, Sep.). The evolution of video quality measurement: From psnr to hybrid metrics. *IEEE Transactions on Broadcasting*, *54*(3), 660-668. doi: 10.1109/TBC.2008.2000733

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), 37–52.

Wolski, K., Giunchi, D., Ye, N., Didyk, P., Myszkowski, K., Mantiuk, R., Seidel, H.-P., … Mantiuk, R. K. (2018, November). Dataset and metrics for predicting local visible differences. *ACM Trans. Graph.*, *37*(5), 172:1–172:14. Retrieved from `http://doi.acm.org/10.1145/3196493` doi: 10.1145/3196493

Wu, J., Shen, X., Zhu, W., & Liu, L. (2013, September). Mesh saliency with global rarity. *Graph. Models*, *75*(5), 255–264. Retrieved from `http://dx.doi.org/10.1016/j.gmod.2013.05.002` doi: 10.1016/j.gmod.2013.05.002

XfrogInc. (2019). Retrieved from `http://xfrog.com/`

Xu, K., Kim, V. G., Huang, Q., Mitra, N., & Kalogerakis, E. (2016). Data-driven shape analysis and processing. In *Siggraph asia 2016 courses* (p. 4).

Ye, P., Kumar, J., & Doermann, D. (2014, June). Beyond human opinion scores: Blind image quality assessment based on synthetic scores. In *2014 ieee conference on computer vision and pattern recognition* (p. 4241-4248). doi: 10.1109/CVPR.2014.540

Zhang, H., & Dana, K. (2018, September). Multi-style generative network for real-time transfer. In *The european conference on computer vision (eccv) workshops.*

Zheng, Y., Zhong, G., Liu, J., Cai, X., & Dong, J. (2014). Visual texture perception with feature learning models and deep architectures. In *Chinese conference on pattern recognition* (pp. 401–410).

Zhou, H., Sun, J., Turk, G., & Rehg, J. M. (2007). Terrain synthesis from digital elevation models. *Transactions on Visualization and Computer Graphics*, *13*(4), 834–848.

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer vision (iccv), 2017 ieee international conference on.*

Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., & Shechtman, E. (2017). Toward multimodal image-to-image translation. In *Advances in neural information processing systems.*

APPENDIX A: IRB CERTIFICATION

PURDUE UNIVERSITY

HUMAN RESEARCH PROTECTION PROGRAM
INSTITUTIONAL REVIEW BOARDS

To:	BENES, BEDRICH RAJASEKARAN, SUREN DEEPAK
From:	DICLEMENTI, JEANNIE D, Chair Social Science IRB
Date:	09/18/2017
Committee Action:(2)	Determined Exempt, Category **(2)**
IRB Action Date:	09 / 18 / 2017
IRB Protocol #:	1709019677
Study Title:	Perceptual Evaluation of Procedurally Modeled Natural Objects

The Institutional Review Board (IRB) has reviewed the above-referenced study application and has determined that it meets the criteria for exemption under 45 CFR 46.101(b).

Before making changes to the study procedures, please submit an Amendment to ensure that the regulatory status of the study has not changed. Changes in key research personnel should also be submitted to the IRB through an amendment.

Refer to our guidance **"Changes Not Requiring Review"** located on our website at http://www.irb/purdue.edu/policies.php. For changes requiring IRB review, please **Create a New Amendment** through the CoeusLite Online Submission System. Please contact our office if you have any questions.

Below is a list of best practices that we request you use when conducting your research. The list contains both general items as well as those specific to the different exemption categories.

General

- To recruit from Purdue University classrooms, the instructor and all others associated with conduct of the course (e.g., teaching assistants) must not be present during announcement of the research opportunity or any recruitment activity. This may be accomplished by announcing, in advance, that class will either start later than usual or end earlier than usual so this activity may occur. It should be emphasized that attendance at the announcement and recruitment are voluntary and the student's attendance and enrollment decision will not be shared with those administering the course.
- If students earn extra credit towards their course grade through participation in a research project conducted by someone other than the course instructor(s), such as in the example above, the students participation should only be shared with the course instructor(s) at the end of the semester. Additionally, instructors who allow extra credit to be earned through participation in research must also provide an opportunity for students to earn comparable extra credit through a non-research activity requiring an amount of time and effort comparable to the research option.

APPENDIX B: IRB AMENDMENT

PURDUE
UNIVERSITY

HUMAN RESEARCH PROTECTION PROGRAM
INSTITUTIONAL REVIEW BOARDS

To:	BEDRICH BENES KNOY 313
From:	JEANNIE DICLEMENTI, Chair Social Science IRB
Date:	07/25/2018
Committee Action:	Amended Exemption Granted
Action Date:	07/24/2018
Protocol Number:	1709019677
Study Title:	Perceptual Evaluation of Procedurally Modeled Natural Objects

The Institutional Review Board (IRB) has reviewed the above-referenced amended project and has determined that it remains exempt. Before making changes to the study procedures, please submit an Amendment to ensure that the regulatory status of the study has not changed. Changes in key research personnel should also be submitted to the IRB through an amendment.

Please retain a copy of this letter for your regulatory records. We appreciate your commitment towards ensuring the ethical conduct of human subject research and wish you well with this study.

Ernest C. Young Hall, 10th Floor - 155 S. Grant St. - West Lafayette, IN 47907-2114 - (765) 494-5942 - Fax: (765) 494-9911

VITA

Suren Deepak Rajasekaran is an alumni from Purdue University (West Lafayette, USA) and High-Performance Computer Graphics Lab (Department of Computer Graphics Technology) with a Ph.D. in Technology (Specializing in Computer Graphics, Computational Science and Engineering (CSE)). Previously, Suren graduated with a Master degree in Computer Graphics that focused on Animation and Simulation topics from Purdue University (West Lafayette, USA) and a Bachelor's degree in Computer Science & Engineering from Anna University (Chennai, India). During the course of his graduate education, he was a course instructor in the Department of Computer Graphics Technology for foundational computer graphics courses, CGT-101 (Foundations of Computer Graphics Technology) and CGT-118 (Fundamentals of Imaging Technology). He has completed internships at Sony US Research Center, Samsung Research America (SRA) and BotVFX in the areas of Human Simulation, GPU Research, and Visual Effects respectively. He had also won multiple awards and grants for research, teaching and leadership services.