Special Section on SBGames 2018

Deep spherical harmonics light probe estimator for mixed reality games

Bruno Augusto Dorta Marques*, Esteban Walter Gonzalez Clua, Cristina Nader Vasconcelos

Instituto de Computação - UFF, Av. Gal. Milton Tavares de Souza, Niterói, RJ, Brazil

ARTICLE INFO

ABSTRACT

The recent developments in virtual and mixed reality by the video game and entertainment industries are responsible for increasing user's visual immersion and provide a better user experience in games and other interactive simulations. However, the interaction between the user and simulated environment still relies on game controllers or other unnatural handheld devices. In the mixed reality context, the usage of more natural and immersive alternative to the game controllers, such as the user's hands, may drastically increase the game interface experience, allowing a personalized visual feedback of the user's interactions in the real-time simulation. There are basically two approaches for including the user's hand: a 3D reconstruction based method, typically based on depth cameras, or an image-based approach, composing the virtual scene with the real images of the user's hands. In the composition of the user's hands and virtual elements, perceptual discrepancies in the illumination of objects may occur, generating an inconsistency in the illumination of the mixed reality environment. A consistent illumination of the environment greatly improves the user's immersion in the mixed reality application. One way to ensure consistent illumination is by estimating the real-world illumination and use this information to adapt the virtual world lighting setting. We present the Spherical Harmonics Light Probe Estimator, a deep learning based technique that estimates the lighting setting of the real-world environment. The method uses a single RGB image and does not requires prior knowledge of the scene. The estimator outputs a light probe of the real-world lighting, represented by 9 spherical harmonics coefficients. The estimated light probe is used to create a composite image containing both real and virtual elements in an environment with a consistent illumination. We validate the technique through synthetic tests achieving an RMS error of 0.0573. We show the usage of the method in an augmented virtuality application.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Mixed Reality (MR) is the mixture of virtual reality environment and the real world. Several applications can benefit from mixed reality because of the increased user's immersion in the simulated environment. This increased immersion can be achieved due to improvements in the realism of the simulated environment and the creation of a personalized experience.

The mixed reality spectrum defines a range of environments between a complete real scene and a totally virtual environment. This term is defined in the reality-virtuality continuum by Milgram et al. [1]. In this spectrum, it is possible to define the Augmented Reality (AR) environment, where most of the environment is composed of objects from the real world. Virtual objects are inserted in the real environment, allowing some kind of interaction with them.

At the other end of the mixed reality spectrum, there is the Augmented Virtuality (AV), where most of the environment is composed of virtual objects. Real objects are inserted in this virtual environment and can interact with the users and the virtual world. An additional form of augmented virtuality is the usage of real-world information, such as movement sensors, GPS location, and weather information.

In the extrema of the reality-virtuality continuum, there are the real environments consisting exclusively of real objects. In the other end of the continuum, there are the virtual environments consisting exclusively of virtual objects, being these virtual environments the essential part of the Virtual Reality (VR) applications.

AR and VR applications usually make use of the user's information, such as the user's location, movements, and image to bring a personalized experience to the users. The personalized experience has been explored by the entertainment industry to attract new users to video-games (Nintendo Wii®, Microsoft Kinect®) and im-

* Corresponding author.
E-mail addresses: brunodorta@id.uff.br (B.A.D. Marques), esteban@ic.uff.br (E.W.G. Clua), crisnv@ic.uff.br (C.N. Vasconcelos).

prove the user's interaction in home cinema. The advances in MR and VR technology have a huge potential to increase even more the personalization of user's experience by introducing the user as an active character in the simulation.

Recent developments in virtual reality technologies are making low-cost Head-Mounted Displays (HMDs) available to the public. The HMDs manufacturers are investing significant technical and monetary resources to create and distribute immersive content for home entertainment, including video games and cinema. The increasing popularity of VR and MR is reflected by the worldwide revenues for mixed reality and virtual reality market that are expected to grow from US$5.2 billion in 2016 to more than US$162 billion in 2020 [2].

1.1. User's immersion

Video game and other simulation contents for VR and MR treat the user as the main character of the retreated history. The usual representation of the user is a virtual character or an animated 3D model, typically named as the avatar.

For VR and MR applications, the avatar is almost exclusively seen from the first-person point of view. Meaning that a camera is positioned in the eyes of the avatar, and what is seen by the camera is the representation of what the avatar is observing in the environment. The first-person point of view also means that for most of the time, the only visible part of the avatar is the upper limbs (hands and arms).

A big improvement in the user's immersion in MR applications can be achieved by substitution of a 3D avatar's reconstruction of the upper limbs by real images of the user's body, captured by a camera positioned at the HMD. This substitution would mean that the user can see the exact real appearance and movement of his arms and hands, including skin color, geometry, and lighting conditions in the mixed reality simulation, increasing dramatically the user's personalization of the experience.

This substitution of the avatar's synthetic upper limbs to the real user's upper limbs is not straightforward. The projection of a real-world 2D footage containing the user's body, captured from a color camera, into the mixed reality environment is a possible approach. The captured footage containing the user's upper limb is projected into the virtual camera plane, being necessary to preprocess the captured footage to remove unwanted objects from the real scene environment.

This approach is capable of accurately representing the user's physical attributes in the virtual environment but fails to merge the real world and virtual world appearance due to the different lighting condition. We call this difference in the lighting condition between the real and virtual environment as the illumination mismatch problem.

1.2. Illumination mismatch

The illumination mismatch problem can affect the way that the user perceives the scene, due to the distinct lighting conditions in the real objects and the virtual scene. This may cause the user perception to loses the sensation of belonging to the scene, leading to a decrease in his self-presence sensation. The illumination mismatch problem can be solved by adjusting the illumination of the real, or virtual, or both environment.

The illumination of the virtual world is known a priori in the virtual environments. The information of the light sources position and their properties are required to correctly render the virtual environment. Since every light source is well known, adjusting their properties is also straightforward.

The illumination in the real environment is unknown to the simulation. There is no readily available information about lighting condition such as light source position, direction, intensity or color in the real environment. This may be particularly critical in cases of dynamic changes of the lighting conditions.

In a typical mixed reality scenario, the knowledge of the user's real environment, including the lighting conditions, needs to be extracted from images captured by an RGB camera. This extraction is not straightforward. Likewise, changing the lighting configuration of a color image with unknown illumination is also a challenging task because the image does not provide explicit information about the geometry of the scene.

1.3. Deep lighting estimation

One important step to solve the illumination mismatch problem is to recognize or estimate the lighting condition of the real environment. Based on this information, it is possible to match the virtual and real environment lighting by changing the lighting setup of the virtual environment, with an adequate level design project.

The matching of lighting conditions in both real and virtual environment is beneficial to the mixed reality spectrum on both ends. On the augmented reality end, virtual objects inserted in the real scenario can have a realistic appearance and behave like a real object. On the augmented virtuality end, a real object can seamlessly be inserted in the virtual environment.

In a practical mixed reality application, the lighting estimation process can't be onerous for the user. Hence, it should not require complicated setup procedures or additional hardware. By the nature of such applications, which implies the user's movement in an interactive simulation, the lighting estimation process should be computationally fast enough to achieve interactive rate and recognize changes in the illumination.

The lighting estimation is a pattern recognition task that can be treated by machine learning algorithms. Among machine learning approaches that could tackle this task, deep learning algorithms [3] have been responsible for most of the success on techniques to classify or recognize patterns in images and videos.

Artificial Neural Networks (ANN) [4] are specialized algorithms for pattern recognition. These algorithms are inspired by the physiological structure of the human brain, where a pattern is learned by a complex connection between cells called neurons. In the computational neural network, data processing cells called artificial neurons are connected in layers. In the past, ANN algorithms made use of architectures containing few layers of neurons.

Advances in processing power, in particular by the development of Graphics Processing Units (GPUs) and the high availability of data have driven the ANN researchers to increase the number of layers in the ANN architectures. Deep learning is a concept that defines techniques whose an architecture of artificial neural network with multiple hidden layers is used to solve a machine learning task.

Given the nature of the lighting estimation problem as a pattern recognition task, we explore the deep learning algorithms to solve the lighting estimation problem. We developed a strategy to estimates the lighting in the real environment based on a color image. The key contributions of this paper are:

- Novelty strategy based on deep learning to estimate lighting from a raw image. This method is particularly more convenient than others since it does not require special devices such as depth cameras, fish eyes lenses or passive probes inserted in the scene. Furthermore, the method does not require previous knowledge of the scene's geometry;
- A method for lighting estimation that is suitable for indoor and outdoor environments;
- Fast inference for interactive environments. While training a deep learning based model can be computationally intensive,