COMPUTER GRAPHICS I COMP.5460

LITERATURE REVIEW 1

Avatar Digitization from a Single Image for Real-Time Rendering

As the title of the research paper states, the research aims at building a system for make 3D avatars out of single images. A fully automatic framework that digitizes a complete 3D head with hair from a single unconstrained image is presented. The system offers a practical and consumer-friendly end-to-end solution for avatar personalization in gaming and social VR applications.. With the involvement of VR and its subsidiaries in emerging technologies and entertainment applications, it has been highlighted how valuable and captivating the immersion of alternate universe can be. It has the potential to revolutionize 3D communication. However, despite efforts in real-time simulation, strand-based representations are still very difficult to integrate into game environments due to their rendering and simulation complexity.

Thus, in this paper the first automatic framework that generates a complete 3D avatar from a single unconstrained image, using high-quality optimized polygonal strips (polystrips or poly cards) for real-time hair rendering is presented. The digitized models are fully rigged with intuitive animation controls such as blendshapes and joint-based skeletons, and can be readily integrated into existing game engines. Subsequently, a complete high-quality facial texture is synthesized using a deep learning-based inference technique. The focus is on a method that produces highly efficient polystrips rather than strands by introducing an automatic hair digitization pipeline for modelling polystrip-based hairstyles. Deep-learning is used for this.

Avatar Modelling Framework: An initial pre-processing step computes pixel-level segmentation of the face and hair regions. The digitization of face and hair since they span entirely different spaces for shape, appearance, and deformation are decoupled. It has the following steps:

1. Image Pre-Processing: The researchers adopted the real-time and automatic semantic segmentation technique which uses a two-stream deconvolution network to predict face and hair regions. This technique produces accurate and robust pixel-level segmentations for unconstrained photographs.
2. Face Digitization: A visibility constraint is incorporated into the model fitting process to improve occlusion handling and non-visible regions.

3. Hair Digitization: A state-of-the-art deep convolutional neural network based on residual learning to extract semantic hair attributes such as hair length, level of baldness, and the existence of hairlines and fringes are used. These hair attributes are compared with a large hairstyle database containing artist created hair polystrip models.
4. Rigging and Animation: They implemented our rig in both the animation tool, Autodesk Maya, and the real-time game engine, Unity and could rig the hair model directly with the skeleton joints of the head in order to add a minimal amount of dynamics for simple head rotations.

Face Digitization: They build a fully textured head model using a multi-linear PCA face model. The extracted high-frequency texture is incomplete from a single-view, and thus the complete texture map using a facial appearance inference method based on deep neural networks was inferred.

1. The first step in this is the 3D Head modelling. The pipeline which is based on morph able face models extended with a PCA-based facial expression model and an efficient optimization based on pixel colour constraints is adopted.
2. The second step is Face Texture Reconstruction. In the above step a low frequency map and a partially visible fine scale structure is obtained. Work has to been done on that to obtain high frequency texture map using deep learning based transfer technique.
3. The third step is developing secondary components. To enhance the realism of the reconstructed avatar, we insert template models for eyes, teeth, gums, and tongue into the reconstructed head model. The reconstructed face model is rescaled and translated to fit a standardized pair of eye balls so that each avatar is aligned as to avoid scale ambiguity during the single-view reconstruction.

Hair Digitization: The following steps lead to complete hair digitization:

1. Hairstyle database: It refers to maintaining a large database of hairstyles of the world.
2. Hair attribute classification: They used 40K images from the CelebA dataset with various hairstyles and collect their hair attributes using AMT.
3. Hairstyle matching: For each hairstyle in the database, pre-render the mask and the orientation map as thumbnails from 35 different views, where 7 angles are uniformly sampled in $[-\pi/4, \pi/4]$ as yaw and 5 angles in $[-\pi/4, \pi/4]$ as pitch. If the hair segmentation mask has multiple connected components due to occlusion or if the hair is partially cropped, then the segmentation descriptor may not be reliable; in this case, we find the most similar hairstyle using the classifiers.
4. Hair mesh fitting: perform spatial deformation in order to fit the hair model to the personalized head model, using an as-rigid-as-possible graph-based deformation model.

5. Hair rendering and texturing: textures are manually designed by an artist based on pre-categorized images that are also used for training. As demonstrated in many games, these type of hair textures can represent a wide range of hair appearances. As different hair types are associated with custom shaders, some styles may be associated with a bump map, which is also prepared by the artist.

Results: A fully-rigged 3D avatars with challenging hairstyles and secondary components for a diverse set of inputs from a wide range of image sets is created. short and long hairstyles of different local structures including straight, wavy, and dreadlock styles were also processed.

High-Quality Single-Shot Capture of Facial Geometry

This paper is about a passive stereo system for capturing the 3D geometry of a face in a single-shot under standard light sources. The system is low-cost and easy to deploy. Results are sub millimetre accurate and commensurate with those from state-of the- art systems based on active lighting, and the models meet the quality requirements of a demanding domain like the movie industry. The main focus is on modification of standard stereo refinement methods to capture pore-scale geometry, using a qualitative approach that produces visually realistic results. These include capture in a studio setup, capture off a consumer binocular-stereo camera, scanning of faces of varying gender and ethnicity and age, capture of highly-transient facial expression, and scanning a physical mask to provide ground-truth validation. This paper presents a passive stereo vision system that computes the 3D geometry of the face with reliability and accuracy on a par with a laser scanner or a structured light system; introduced an image-based embossing technique to capture mesoscopic facial geometry1, so that the quality of synthesized faces from our system equals that achieved with gradient-based illumination.

Face scanning is an important sub task. It has three sections:

Calibration: The method requires a small number of views, typically one to three views, of a sphere augmented with fiducials. Each fiducial is a double circle. it is suited to face capture because a calibration sphere that is approximately head-sized and placed at the intended position of the subject is therefore well-placed for the cameras. Unlike a calibration plane, a sphere has no preferred direction in space, making it appropriate for a setup in which circum positioned cameras are directed inward towards an object of interest. Unlike an LED-based calibration, the method requires only a small number of views and provides sub-pixel accurate features.

Pairwise Stereo-reconstruction: First, the face is segmented out of the images using cues of background subtraction and skin color. Matching is done pairwise between neighboring cameras2, and at pixel level to establish dense matches across the face. The image resolution at the lowest-resolution layer of the pyramid is chosen to be around 150×150 pixels, but this is approximate and the criteria is simply that the major facial features are still visible.

Refinement: This section describes the refinement method that was utilized in Section 2. The refinement consists of a linear combination of two terms: a photometric consistency term dp that favors solutions with high NCC and a surface consistency term ds that favors smooth solutions. It has three sub sections:

1. Disparity Map Refinement
2. Surface Refinement
3. Mesoscopic Augmentation: The refinement in Surface Refinemnet results in surface geometry that is smooth across skin pores and fine wrinkles, because the disparity change across such a feature is too small to detect. The result is flatness and lack of realism in synthesized views of the face. On the other hand, visual inspection shows the obvious presence of pores and fine wrinkles in the images. This is due to the fact that light reflected by a diffuse surface is related to the integral of the incoming light.

Results: contrast, passive stereo vision uses single-shot capture under standard light sources. And commodity cameras now routinely have the image resolution to reveal individual skin pores, so that faces provide the kind of dense evenly-distributed texture that is perfect for stereo matching and 3D reconstruction.

Similarity and differences between the two papers

The first paper talks about developing a system for digitization of the whole face for real time. They want the system to be fully automatic. It has taken into consideration many aspects from face to hair modelling. The second paper also aims at creating high quality 3D face from a single shot. Having said that, it can be said that both paper want to achieve more or less the same objective. The difference between them is the methods used and the final results. The paper has used manual help like placing 6 cameras at different angles and capturing from it. The first paper aims at creating a animated view of the face, not just the 3D face like the second one.

References:

Avatar Digitization from a Single Image for Real-Time Rendering

ACM Transactions on Graphics, Vol. 36, No. 6, Article 195. Publication date: November 2017

High-Quality Single-Shot Capture of Facial Geometry

ACM Transactions on Graphics, Vol. 29, No. 4, Article 40, Publication date: July 2010