Diss. ETH No. 23032

Geometric Methods for Realistic Animation of Faces

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by
Amit H. Bermano
MSc in Computer Science, The Technion - Israel Institue of Technology, Israel
born 20.11.1980
citizen of Israel

accepted on the recommendation of
Prof. Dr. Markus Gross, examiner
Prof. Dr. Olga Sorkine-Hornung, co-examiner
Prof. Dr. Bernd Bickel, co-examiner

2015

Abstract

Realistic facial synthesis is one of the most fundamental problems in computer graphics, and has been sought after for approximately four decades. It is desired in a wide variety of fields, such as character animation for films and advertising, computer games, video teleconferencing, user-interface agents and avatars, and facial surgery planning. Humans, on the other hand, are experts in identifying every detail and every regularity or variation in proportion from one individual to the next. The task of creating a realistic human face is elusive due to this, as well as many other factors. Among which are complex surface details, spatially and temporally varying skin texture and subtle emotions that are conveyed through even more subtle motions.

In this thesis, we present the most commonly practiced facial content creation process, and contribute to the quality of each of its steps. The proposed algorithms significantly increase the level of realism attained by each step and therefore substantially reduce the amount of manual labor required for production quality facial content. The thesis contains three parts, each contributing to one step of the facial content creation pipeline.

In the first part, we aim at greatly increasing the fidelity of facial performance captures, and present the first method for detailed spatio-temporal reconstruction of eyelids. Easily integrable with existing high quality facial performance capture approaches, this method generates a person-specific, time-varying eyelid reconstruction with anatomically plausible deformations. Our approach is to combine a geometric deformation model with image data, leveraging multi-view stereo, optical flow, contour tracking and wrinkle detection from local skin appearance. Our deformation model serves as a prior that enables reconstruction of eyelids even under strong self-occlusions caused by rolling and folding skin as the eye opens and closes.

In the second part, we contribute to the authoring step of the creation process. We present a method for adding fine-scale details and expressiveness to low-resolution art-directed facial performances. Employing a high-resolution facial performance capture system, we augment artist friendly content, such as those created manually using a rig, via marker-based capture, by fitting a morphable model to a video, or through Kinect-based reconstruction. From the high fidelity

captured data, our system encodes subtle spatial and temporal deformation details specific to that particular individual, and composes the relevant ones to the desired input animation. The resulting animations exhibit compelling animations with nuances and fine spatial details that match captured performances, while preserving the artistic intent authored by the low-resolution input sequences, outperforming current state-of-the-art in example-based facial animation.

The third part of the dissertation proposes to enrich digital facial content by adding a significant sense of presence. Replacing the classic 2D or 3D displaying techniques of digital content, we propose the first complete process for augmenting deforming physical avatars using projector-based illumination. Physical avatars have been long used to give physical presence to a character, both in the field of entertainment and teleconferencing. Using a human-shaped display surface provides depth cues and multiple observers with their own perspectives. Such physical avatars, however, suffer from limited movement and expressiveness due to mechanical constraints. Given an input animation, our system decomposes the motion into low-frequency motion that can be physically reproduced by a robotic head and high-frequency details that are added using projected shading. The result of our system is a highly expressive physical avatar that features facial details and motion otherwise unattainable due to physical constraints.

Zusammenfassung

Realistische Gesichtssynthese ist eines der fundamentalen Probleme in der Computer Graphik, woran seit Jahrzehnten gearbeitet wird. Gesichtssynthese wird in verschiedenen Gebieten angewendet, zum Beispiel in der Charakteranimation für Filme und Werbung, Computer Spiele, Video Telekonferenzen, User-Interface Agenten und Avatare und Gesichtsoperationsplanung. Menschen sind Experten darin, jedes Detail und jede Irregularität in den Proportionen des Gesichts zu erkennen. Somit ist die Aufgabe, ein realistische Gesicht zu kreieren, sehr schwierig. Dazu gehören komplexe Oberflächendetails, räumlich und zeitlich sich verändernde Hauttexturen und subtile Emotionen, die durch sogar noch subtilere Bewegungen übermittelt werden. In dieser Arbeit präsentieren wir den gebräuchlichsten Prozess für das Kreieren von Gesichtern und leisten einen Beitrag zur Qualität für jeden Schritt. Die vorgeschlagenen Algorithmen verbessern den Realismus in jedem Schritt erheblich und verringern somit die Handarbeit, die nötig ist, um Produktionsqualität zu erreichen. Die Arbeit ist in drei Teile aufgeteilt, wobei jeder Teil zu einem Schritt im Gesichtskreationsprozess beiträgt.

Der erste Teil zielt darauf, die Genauigkeit von Gesichtsdarbietungen erheblich zu verbessern und präsentiert die erste Methode für eine detaillierte räumlich-zeitliche Rekonstruktion von Augenlidern. Die Methode ist einfach integrierbar in existierende Ansätzen für hochqualitative Gesichtsdarbietungen. Sie generiert eine personen-spezifische und zeitabhängige Rekonstruktion von Augenlidern mit anatomisch plausiblen Deformationen. Unser Vorgehen besteht darin, ein geometrisches Deformationsmodell mit Bilderdaten zu kombinieren und dabei die Stereoperspektive, optischer Fluss, Kantenverfolgung und Faltenerkennung auszunutzen. Unser Deformationsmodell ermöglicht uns die Rekonstruktion von Augenlidern auch mit starken Selbstokklusionen, die durch rollende und faltende Haut während das Auge sich öffnet und schliesst, entstehen.

Im zweiten Teil tragen wir zum verfassenden Schritt im Gesichtskreationsprozess bei. Wir präsentieren eine Methode um feine Details und Ausdrucksfähigkeit zu handgemachten Gesichtsdarbietungen mit nur tiefer Auflösung hinzuzufügen. Wir benutzen ein hochauflösendes Gesichtsdarbietungsaufnahmesystem um künstlerische Inhalte zu ergänzen. Das System unterstützt Inhalte, die von Hand gemacht wurden mittels eines Rigs, markerbasierende Aufnahmen, durch einpassen eines wandelbaren Modells von einem Video erstellte Inhalte und Kinect-

basierte Rekonstruktionen. Von den hochauflösenden Aufnahmedaten enkodiert unser System subtile räumliche und zeitliche Deformationsdetails für das spezifische Individuum. Die resultierenden Animationen zeigen überzeugende Animationen mit Nuancen und feinen Details, die den Bewegungsaufnahmen entsprechen. Dabei bleibt die künstlerische Absicht der tiefauflösenden Eingangssequenzen erhalten. Sie übertreffen die zur Zeit modernsten beispielbasierten Gesichtsanimationen.

Der dritte Teil der Dissertation befasst sich mit dem Bereichern von digitalen Gesichtern durch hinzufügen eines signifikanten Präsenzgefühlts. Wir beabsichtigen 2D oder 3D Bildschirmtechnologien für digitalen Inhalt mit einem Prozess zur Augmentation von deformierbaren physischen Avataren mit projektorbasierter Illumination zu ersetzen. Physische Avatare werden schon lange verwendet, um einem Wesen physische Präsenz zu verleihen, in der Unterhaltung sowie für Telekonferenzen. Eine menschlich geformte Anzeige birgt Tiefenanhaltspunkte und bietet mehreren Beobachtern eine eigene Perspektive. Jedoch leiden solche physische Avatare an eingeschränkter Bewegungsfreiheit und Ausdrucksfähigkeit aufgrund mechanischer Beschränkungen. Unser System zerlegt die Eingangsanimation in tieffrequente Bewegungen, welche ein Roboterkopf ausführen kann, und hochfrequente Details, welche mit Projektion hinzugefügt werden. Das Resultat ist ein hoch ausdrucksfähiger physischer Avatar, der Gesichtsdetails sowie Bewegungen umfasst, die unerreichbar wären mit einem rein mechanischen Avatar.