

DISS. ETH NO. 19517

Capturing and Synthesizing Hand-Object Interaction

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences (Dr. sc. ETH Zürich)

presented by

Henning Hamer

Dipl.-Inf., University of Ulm
born August 10, 1979
citizen of Konstanz, Germany

accepted on the recommendation of

Prof. Dr. Luc Van Gool, examiner
Prof. Dr. Reinhard Klein, co-examiner
Prof. Dr. Konrad Schindler, co-examiner

2011

Abstract

Humans use their hands for interacting with their environment. In particular the ability of hands to manipulate objects dominates our daily life. As a consequence, there are many useful digital applications in the context of hand-object interaction. Realizing such applications requires both knowledge about the manipulating hand and knowledge regarding the manipulated object. The proposed thesis is concerned with the extraction of such information by means of computer vision methods. It also explores interdependencies between object and hand: the manipulation of a specific object implicates appropriate executing hand poses. This ultimately leads to a method to synthesize static grasps as well as dynamic hand motion for object manipulation: the knowledge gained with a camera system is used to ease the generation process of computer animations containing hand-object interaction.

To gather information about the manipulating hand, we first develop a markerless method for tracking the articulated pose of a hand interacting with an object. The scenario implies the challenge of severe self-occlusions of the hand as well as occlusions by the object. We approach the task by belief propagation and use range and color data as input. This data is delivered by a real-time structured-light system.

Next, the manipulated object is not only considered as an occluder but plays an active role. We introduce an object-dependent hand pose prior. The prior represents knowledge regarding the manipulation of a certain object by a certain hand. It can be generalized towards new hands and new objects of the same object class (e.g., from one cup to another). We present two applications of this: improved 3d hand tracking, and grasp synthesis for unobserved hands and objects. Grasp synthesis is highly relevant in the fields of robotics and computer graphics.

Finally, we examine temporal aspects of hand pose generation and extend static grasp synthesis to dynamic hand motion synthesis. The idea is to generate

computer animations containing hand-object interaction solely by animating the object. The corresponding hand information is derived from observation. While animating an object can be easily done by a 3d artist, manually animating a hand is substantially more complicated due to the many degrees of freedom. Our method requires training once for an object of interest. After that, arbitrary manipulation sequences can be realized in 3d modeling software like Autodesk Maya.

Zusammenfassung

Menschen gebrauchen ihre Hände für die Interaktion mit ihrer Umgebung. Besonders die Möglichkeit, mit Händen Objekte zu manipulieren, beherrscht unseren Alltag. Aus dieser Tatsache ergibt sich eine Vielzahl von nützlichen digitalen Anwendungen im Bereich der Hand-Objekt-Interaktion. Die Realisierung solcher Anwendungen erfordert einerseits Wissen über die manipulierende Hand selbst und andererseits Wissen über das manipulierte Objekt. Die vorgelegte Dissertation befasst sich mit der Gewinnung solchen Wissens mittels Computer Vision Methoden. Darüber hinaus werden Abhängigkeiten zwischen dem Objekt und der Hand erforscht: die Manipulation eines speziellen Objekts impliziert entsprechende ausführende Handposen. Dies führt schließlich zu einer Methode zur Erzeugung von sowohl statischen als auch dynamischen Handposen: das Wissen, welches mit einem Kamerasystem gewonnen wird, dient der erleichterten Generierung von Animationen mit Hand-Objekt-Interaktion.

Um Informationen über die manipulierende Hand zu erhalten, entwickeln wir zunächst eine Methode ohne Marker zum Tracken der artikulierten Pose einer Hand, welche mit einem Objekt interagiert. Dieses Szenario bringt Schwierigkeiten mit sich, die durch starke Selbstverdeckungen der Hand und Verdeckungen durch das Objekt bedingt sind. Wir begegnen den Problemen mit dem Belief-Propagation Algorithmus und verwenden sowohl Tiefen- als auch Farbinformationen. Diese Daten werden in Echtzeit von einem Structured-Light-System geliefert.

Danach fungiert das manipulierte Objekt nicht mehr nur als Verdeckter sondern nimmt eine aktive Rolle ein. Wir stellen einen objekt-spezifischen Handposenprior vor. Dieser Prior repräsentiert Wissen bezüglich der Manipulation eines bestimmten Objekts durch eine bestimmte Hand, und er ist generalisierbar bezüglich neuer Hände und neuer Objekte der gleichen Objektklasse (zum Beispiel von einer Tasse zu einer anderen). Wir präsentieren zwei Anwendungen hiervon: erstens verbessertes 3d Handtracking und zweitens Griffsynthese

für zuvor nicht beobachtete Hände und Objekte. Die Griffsynthese hat große Relevanz in den Feldern der Robotik und der Computergraphik.

Schließlich untersuchen wir die zeitlichen Aspekte der Handposenerzeugung und erweitern die statische Griffsynthese hin zur Synthese von dynamischen Handbewegungen. Die Grundidee ist die Erzeugung von Computeranimationen, die Hand-Objekt-Interaktion enthalten, allein durch die Animation des Objekts. Die entsprechende Handinformation wird von Beobachtungen abgeleitet. Während es für einen 3d-Animationskünstler relativ einfach ist, ein Objekt zu animieren, ist die direkte Animation einer Hand wegen der vielen Freiheitsgrade deutlich schwieriger. Unsere Methode benötigt eine einmalige Trainingsprozedur für ein gewünschtes Objekt. Danach können beliebige Manipulationssequenzen in 3d Modellierungssoftware, wie zum Beispiel Autodesk Maya, realisiert werden.