

DISS. ETH NO. 22257

Advanced Energy Optimization Algorithms for 2D and 3D Video-Based Graphics Applications

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

Manuel Johannes Lang
MSc ETH CS, ETH Zurich, Switzerland
born on 16.10.1982
citizen of Austria

accepted on the recommendation of

Prof. Dr. Markus Gross, examiner
Dr. Aljoša Smolić, co-examiner
Prof. Dr. Marc Pollefeys, co-examiner

2014

Abstract

Videos have become an important part in our daily lives. Thankfully, the recent shift to an entire digital video pipeline allowed the introduction of improved tools and algorithms for professional video processing. Therefore, despite its complexity, video processing has never been easier to use and more available to the general public than it is today.

However, due to the rapid technological advances new challenges in video processing have emerged. The ability to capture high definition content has often outpaced the progress of computing hardware. Furthermore, with the renaissance of stereoscopic 3D a new dimension is added to video processing algorithms. This additionally increases computational demands as well as it introduces its own sets of problems. Moreover, the ubiquitous of currently available displays results in a wide range of possible video output formats. Viewers consume video content on small handhelds as well as on high definition glasses-free 3D displays. This requires sophisticated video processing methods to ensure best viewing conditions on all output devices.

This thesis aims at improving the way 2D and 3D video content can be adapted and artistically enhanced. The first two challenges emerge from the content-display gap. Often the resolution and aspect ratio of the output display may significantly vary from the original content. This thesis introduces a novel image domain warping (IDW) framework and shows that it can be utilized to fit content to different display dimensions without introducing too much noticeable artifacts. We present a novel efficient system that gives complete artistic control over this retargeting process.

A similar content-display gap also exists for stereoscopic productions. Different display dimensions result in different depth perception. To counteract those undesired depth changes we will extend the technique of IDW to be able to change the depth perception of 3D content. Our novel method does not only allow to close the 3D-content display gap, but also gives artists a new innovative post processing tool to locally modify the depth composition. We introduce disparity mapping operators. These operators formalize insights from perception and production rules and are the basis for a general framework for stereoscopic disparity editing. Moreover, we use the same novel IDW approach to generate multi-view video

from stereoscopic input. This is an essential tool to be able to consume stereoscopic content on advanced glasses-free auto-stereoscopic displays.

One other energy optimization algorithm presented in this thesis tackles the problem of information propagation spatially and temporally in a video sequence. To ensure consistent and content-aware propagation such algorithms usually require complex dependencies and therefore result in a heavy and difficult optimization problems. The general processing framework, which this thesis introduces however, is optimized to still work on high definition content. Due to the reduction of per-pixel complexity and the increase of the amount of information that can be considered simultaneously it solves efficiently and accurately many data propagation problems. We show that our novel video processing framework can significantly improve and accelerate various video-processing tasks including such important tasks as optical flow computation, disparity estimation, depth up sampling, colorization, and saliency computation.

Zusammenfassung

Der Konsum von Videoinhalten wurde in den letzten Jahren zu einem wichtigen Teil unseres täglichen Lebens. Die jüngsten Entwicklungen hin zu einer vollkommen digitalen Video-Pipeline erlaubten es bessere Werkzeuge und Algorithmen für die professionelle Videoverarbeitung einzuführen. Dadurch ist die Videoverarbeitung, trotz ihrer Komplexität, einfacher und weiter verbreitet als jemals zuvor.

Aufgrund des rasanten technologischen Fortschritts sind neue Herausforderungen in der Videoverarbeitung entstanden. Die Fähigkeit immer bessere, hochaufgelöste Inhalte aufzuzeichnen überholt oft die Weiterentwicklung der zur Verarbeitung verwendeten Computer-Hardware. Zusätzlich, mit der Renaissance des 3D-Films, wurde noch eine weitere Verarbeitungsdimension zu den meisten Videoalgorithmen hinzugefügt. Dies erhöht zusätzlich den Rechenaufwand und erzeugt weiters eine ganze Reihe an speziellen neuen Problemen. Des Weiteren ergeben sich auch durch die weite Verbreitung von Displays mit verschiedensten Auflösungen und Ausgabeformaten neue Herausforderungen für die Videoverarbeitung. Zuschauer konsumieren Videoinhalte sowohl auf kleinen Handhelds als wie auch auf grossen High-Definition 3D-Displays. Um trotzdem auf allen Ausgabegeräten eine optimale Qualität zu erhalten, werden hochentwickelte Anpassungsalgorithmen benötigt.

Diese Arbeit befasst sich mit der Verbesserung der Art und Weise wie 2D- und 3D-Videoinhalte technisch sowie auch künstlerisch angepasst werden können. Als Erstes widmen wir uns dem Problem der Anpassung von vorgegebenen 2D Videos an bestimmte Ausgabegeräte (Content-Display-Gap). Oft unterscheidet sich die Auflösung und das Seitenverhältnis des Ausgabegerätes stark von den Eigenschaften der Aufnahme. Um dieses Problem zu lösen stellt diese Arbeit zuerst ein neuartiges Werkzeug zur Video-Verformung im Bildbereich (image domain warping, IDW) vor. Wir zeigen dann, dass mit diesem IDW-Werkzeug Videos an beliebige Seitenverhältnisse angepasst werden können, ohne dass dabei starke Verzerrungen in wichtigen Bereichen der Videos entstehen. Unser neuartiges und effizientes System gewährt die volle Kontrolle über diese Bildverformung und erlaubt dadurch auch die gewollte künstlerische Originalbildkomposition zu erhalten.

Ein ähnliches Video-Anpassungsproblem gibt es auch im Bereich der stereosko-

pischen 3D Videoproduktion. Verschiedene Anzeigegeräte führen hier oft dazu, dass ein und derselbe Inhalt eine andere Tiefenwahrnehmung erzeugt. Um dieser unerwünschten Tiefenänderungen entgegenzuwirken erweitern wir unser Bildverformungswerkzeug so, dass wir die Tiefenwahrnehmung von beliebigen 3D-Inhalten anpassen können. Dieses neue innovative 3D Bearbeitungswerkzeug ermöglicht nicht nur Inhalte an 3D Ausgabegeräte anzupassen, sondern erlaubt es Videobearbeitern und Videokünstlern auch in der Videonachbearbeitung die Tiefenzusammensetzung einer aufgezeichnete Szene zu ändern. Dazu stellen wir das Konzept von Deviationsanpassungsfunktionen vor. Diese Funktionen formalisieren Erkenntnisse aus der Wahrnehmungslehre sowie allgemeine bekannte Produktionsregeln für 3D Inhalte, und sind damit die Grundlage für einen allgemeinen Ansatz zur nachträglichen Bearbeitung der Deviationen und der damit verbundenen Tiefenzusammensetzung von 3D Videos. Wir zeigen auch dass unser auf reiner zweidimensionaler Videoverformung basierendes 3D Bearbeitungswerkzeug zudem dafür verwendet werden kann zusätzliche neue Ansichten aus anderen Blickwinkeln zu erzeugen. Damit ist es ein wichtiges Instrument um stereoskopische Aufnahmen bestehend aus nur zwei Ansichten an Ausgabegeräte mit vielen Ansichten (autostereoskopische Bildschirme) anzupassen.

Ein weiterer Energieoptimierungs-Algorithmus, der in dieser Arbeit beschrieben wird, befasst sich mit dem Problem der räumlichen und zeitlichen Diffusion von Information in einer Videosequenz. Um sicherzustellen dass Information konsistent und unter Berücksichtigung der Videoinhalte propagiert werden können werden oft Algorithmen mit komplexen Abhängigkeiten zwischen allen Bildern und allen Bildpunkten benutzt. Dies führt zu schwierigen und rechenaufwendigen Optimierungsproblemen. In dieser Arbeit führen wir deshalb eine neuartige Videoverarbeitungsmethode ein die auch noch für hochaufgelöste Videos effizient arbeitet. Indem wir den Aufwand pro Bildpunkt reduzieren und damit erreichen dass wir gleichzeitig mehr Bildinformation berücksichtigen können, gelingt es uns viele Informationsdiffusionsproblem effizient und in hoher Qualität zu lösen. Wir zeigen dass unsere neue Methode für viele wichtige Anwendungen wie zum Beispiel die optische Flussberechnung, die Deviationsschätzung, die Tiefenkartenrekonstruktion sowie die Filmkolorierung effizient verwendet werden kann.