Diss. ETH No. 21116

Hardware Architectures for Real-Time Video Processing and View Synthesis

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

PIERRE GREISEN

Dipl. El.-Ing. ETH,

Ingénieur des Arts et Manufactures, EC Paris

born 26.08.1982

citizen of Luxembourg

accepted on the recommendation of

Prof. Dr. Hubert Kaeslin, examiner

Prof. Dr. Markus Gross, co-examiner

Prof. Dr. Andreas Burg, co-examiner

Dr. Aljosa Smolic, co-examiner

2013

Abstract

Today, we consume video content on a variety of display devices. However, the resulting variety of display formats and standards conflicts with the rigid video acquisition process. Quality video content acquisition is expensive, both financially and technically, and is therefore typically produced for a very specific display format. The objective of this thesis is to close the gap between fixed-format content acquisition and variable-format content display using spatially-varying warping and smart acquisition.

Instead of generating and transmitting video content in various formats, e.g., aspect ratio or number of views, the spatially-varying warping framework transforms the video based on meta-information, scene analysis results, and display requirements. A first part of this thesis is therefore concerned with extracting information from the scene in real-time. In particular, we provide efficient hardware architectures for visual importance estimation and depth from high-definition stereoscopic video. In a second part, we analyze and develop real-time techniques to generate a deformation grid based on scene information and display requirements. To this end, we provide FPGA implementations of sparse solvers for systems of equations in the order of 10^4 to 10^5 unknowns. In a final part, we analyze non-linear, spatially-varying warping and develop an adaptive resampling algorithm as well as several VLSI architectures for warping high-definition and high-quality video in real-time.

The different steps of the spatially-varying warping framework are put together to provide real-time hardware architectures for automatic display adaptation applications such as aspect ratio retargeting or stereoscopic remapping. The efficient algorithms and architectures are

thereby targeted for fixed-function hardware in end-user display devices. Instead of using general-purpose GPUs or CPUs, fixed-function hardware has significant advantages in terms of computational power per Watt and computational power per area, and are hence better suited to be integrated into end-user devices.

Finally, to improve acquisition, we provide a computational stereo camera system that reduces the expenses of acquisition. To this end, the camera system controls the camera settings such as interaxial distance or focus in real-time based on analysis results from the video stream. The resulting system greatly alleviates stereoscopic video acquisition.

Zusammenfassung

Heutzutage erreichen uns Video Inhalte auf einer Vielzahl von Anzeigegeräten. Allerdings steht die resultierende Vielzahl von Anzeigeformaten und Standards im Gegensatz zum eher schwerfälligen Video-Akquisition Prozess. Die Aufnahme von Qualitätsvideo ist kostspielig, sowohl finanziell als auch technisch, und wird daher in der Regel für ein bestimmtes Display-Format produziert. Das Ziel dieser Arbeit ist es, die Lücke zwischen Video-Akquisition mit fixen Einstellungen und Video Darstellung auf variablen Displays zu schließen. Dies wird erreicht mittels räumlich variierenden Bildverzerrungen und intelligenter Video-Akquisition.

Statt Produktion und Übertragung von Video in verschiedenen Formaten, z.B. Seitenverhältnis oder Anzahl der Ansichten, wird das Video mittels räumlich variierenden Bildverzerrungen an die Gegebenheiten des Displays angepasst, und zwar basierend auf Meta-Informationen und automatischer Analyse der Szene. Ein erster Teil dieser Arbeit beschäftigt sich demnach mit dem Extrahieren von Informationen aus der Szene in Echtzeit. Wir zeigen effiziente Hardware-Architekturen für die Schätzung von visuell wichtigen Objekten sowie Tiefenschätzung aus hoch auflösendem stereoskopischem Video. In einem zweiten Teil analysieren und entwickeln wir Lösungen zum Generieren von Deformationsraster basierend auf Information der Szene und Display Anforderungen. Zu diesem Zweck entwickeln wir FPGA-Implementierungen von Solvern für lineare Gleichungssysteme mit 10^4 bis 10^5 Unbekannten. In einem abschließenden Teil analysieren wir nicht-lineare, räumlich variierende Bildverzerrungen mittels der generierten Deformationsraster. Der spezifische Beitrag von diesem Teil der Arbeit besteht im Entwickeln eines adaptiven

Resampling-Algorithmus sowie mehrerer VLSI Architekturen für das Verzerren von hoch auflösendem und qualitativ hochwertigem Video.

Die verschiedenen Schritte hin zum Bildverzerren werden kombiniert um Hardware-Architekturen für automatisierte Display Anpassungen zu demonstrieren. Beispiele hierfuer sind automatisches Anpassen von Seitenverhältnis und stereoskopischem Video. Die effizienten Algorithmen und Architekturen werden entwickelt um auf dedizierter Hardware in Endnutzer-Display-Geräte zu laufen. Anstelle der Verwendung von Universal-Prozessoren wie GPUs oder CPUs, hat dedizierte Hardware erhebliche Vorteile in Bezug auf die Rechenleistung pro Watt und Rechenleistung pro Fläche und ist daher besser geeignet für diese Art von Anwendungen in Endnutzer Geräten.

Schlußendlich, um die Akquisition zu verbessern, entwickeln wir ein Stereokamera-System, das die Kosten der Akquisition reduziert. Zu diesem Zweck steuert das Kamera-System die Kameraeinstellungen wie interaxialer Abstand oder Fokus in Echtzeit basierend auf Anaylseresultate wie Tiefenschätzung. Das resultierende System kann somit erheblich die stereoskopische Video-Akquisition erleichtern.