



FernUniversität in Hagen
Lehrgebiet Mensch-Computer-Interaktion

Abschätzung des Disparitätsraums für das Stereo-Matching mithilfe von Bildmerkmalen

- ▶ Erarbeitet im Rahmen der Masterarbeit (eingereicht Sept. 2012):
„Beleuchtungsinvariante und rauschintensive Disparitätskartenberechnung“
- ▶ Autor: Christoph Drexler, Innsbruck, christoph@drexler.eu
- ▶ Betreuung: Dr. Klaus Häming, Prof. Dr. Gabriele Peters,
Lehrgebiet Mensch-Computer-Interaktion, Fernuniversität in Hagen

Präsentation im Rahmen des Wissenschafts- und Absolventenworkshops
auf den Informatiktagen der Gesellschaft für Informatik 2013-03-22, 11:45 Uhr

Stereo-Matching - „Standardfall“



FernUniversität in Hagen
Lehrgebiet Mensch-Computer-Interaktion

- Off-axis-Konstellation, d.h.:
 - parallele Kameraachsen
 - auf gleicher Höhe
 - in gleichem Abstand zur Szene
- kalibriert
- kontrollierte Beleuchtungsbedingungen
- sehr viele Autor/inn/en entwerfen ihre Algorithmen für diesen „Standardfall“ (s. Middlebury Datasets, Middlebury Evaluation)



Stereo-Matching - „Standardfall“



- Berechnung der Disparität (=Positionsdifferenz) jedes einzelnen Pixels
 - lässt Rückschlüsse auf die Tiefendimension der Szene zu
 - „lokale Algorithmen“: Berechnung der Kosten für das Matching zweier Bildpixel
- Suche eingeschränkt auf selbe Bildzeile
- Disparitätsraum meist vorgegeben (s. Middlebury Evaluation)
- Berechnung vergleichsweise einfach, Bilderzeugung relativ aufwändig (Kalibrierung, Sonderequipment)
- „Standardfall“ ist eigentlich ein sehr spezieller „Sonderfall“



H Disparität



Stereo-Matching – Allgemeiner Fall



FernUniversität in Hagen
Lehrgebiet Mensch-Computer-Interaktion

- Kameraachsen stehen windschief zueinander
- dadurch deutlich größerer Disparitätsraum
 - Off-Axis: von 27 bis 81 = 55 Disparitätsstufen
 - hier gezeigtes Beispiel: von -56 bis 118 = 175 Disparitätsstufen
- Kontext: *Freihand*erfassung von 3D-Objekten
 - ermöglicht kaum Einschränkungen bzgl. Disparitätsraum
 - theoretisch Disparitäten von $-{\text{Bildbreite}}$ bis $+{\text{Bildbreite}}$ möglich



Abschätzung des Disparitätsraums



FernUniversität in Hagen
Lehrgebiet Mensch-Computer-Interaktion

- Down- und Upscaling

- z. B. Reduktion von 800x600 auf 200x150 Pixel

- Berechnung mithilfe der verkleinerten Bilder für den Disparitätsraum $[-100, 100]$

- Berechnung mithilfe der großen Bilder für die Disparitätsstufen $[d_{\min}, d_{\max}]$ oder $[d - \epsilon, d + \epsilon]$, wobei $d := 4 * \text{bei der vorherigen Berechnung erhaltenes Ergebnis}$

- für andere Zwecke bereits erprobtes Verfahren,

- z. B. [Hirschmüller2008] zur effizienten Berechnung der Kostenfunktion HMI

- [Hirschmüller2008]: Hirschmüller, Heiko: Stereo Processing by Semi-Global Matching and Mutual Information. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 30 (2008), S. 328-341. Online: http://elib.dlr.de/55367/1/Stereo_Processing-Hirschmüller.pdf (Abrufdatum: 2013-03-15).

- z. B. [Won2011] zur Verringerung des Speicherbedarfs und effizienten Berechnung

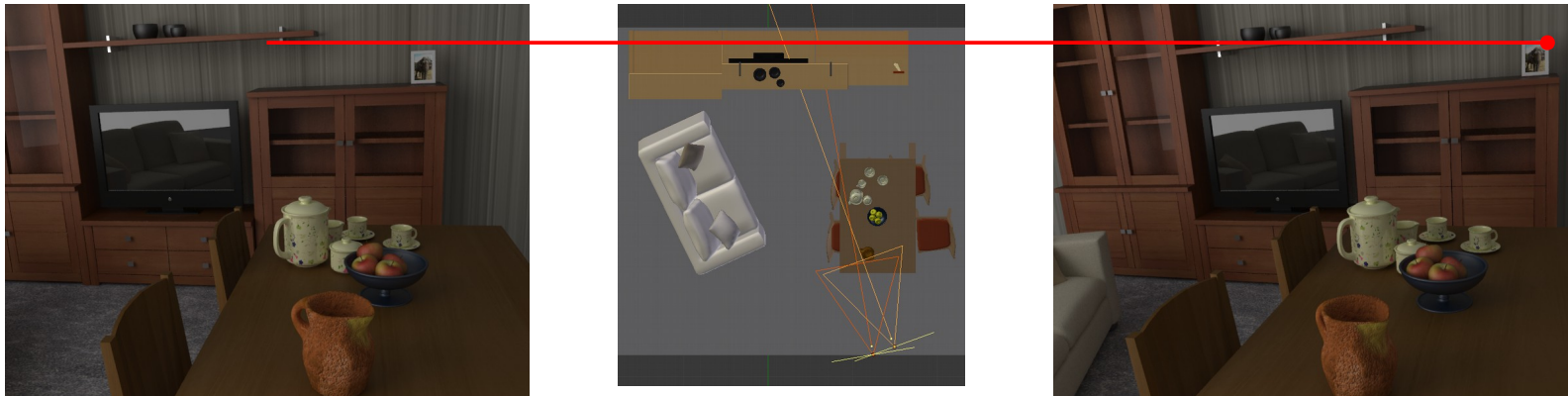
- [Won2011]: Won, Kwang Hee ; Jung, Soon Ki: hSGM: Hierarchical Pyramid Based Stereo Matching Algorithm. In: Blanc-Talon, Jaques et al. (Hrsg.): Advances Concepts for Intelligent Vision Systems. Berlin ; Heidelberg : Springer, 2011 (Lecture Notes in Computer Science ; 6915), S. 693-701. Online: http://dx.doi.org/10.1007/978-3-642-23687-7_62.

- (neue?) Idee: Verwendung der für die Rektifizierung ermittelten Bildmerkmale (Features)

Exkurs: Rektifizierung



- Allgemeiner Fall (windschiefe oder kreuzende Kameraachsen): korrespondierende Bildpunkte liegen nicht notwendigerweise auf derselben Bildzeile



- Durch die sog. Rektifizierung wird das Bild so umgerechnet (verzerrt), dass zusammengehörige Bildpunkte wieder auf einer Bildzeile liegen



Exkurs: Rektifizierung



FernUniversität in Hagen
Lehrgebiet Mensch-Computer-Interaktion

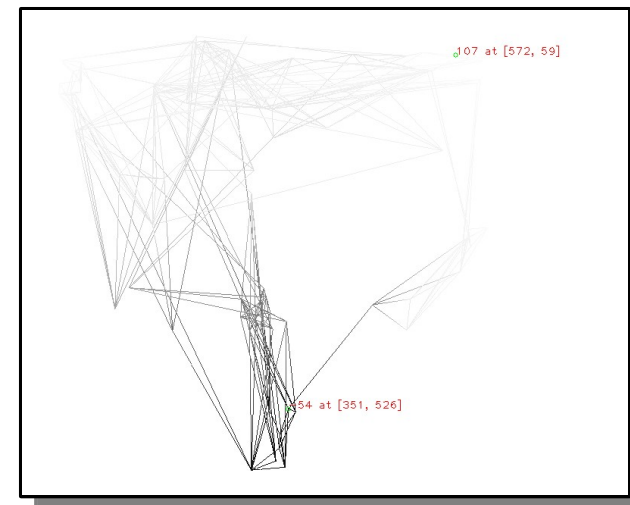
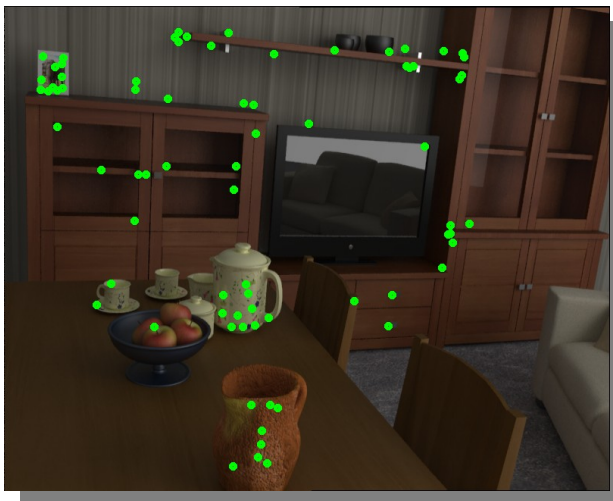
- benötigt mind. 7 Paare von Bildmerkmalen („Features“)
- üblicherweise sehr viel mehr korrespondierende Bildpunkte – im 3-stelligen Bereich
- Disparitäten dieser Bildmerkmale sind bekannt bzw. sehr einfach berechenbar
- Ausgangspunkt zur Abschätzung des Disparitätsraums



Abschätzung des Disparitätsraums



1. Setze den Disparitäts(such)raum für die gefundenen Bildmerkmale auf den errechneten Disparitätswert: $[d, d+1[$
2. Finde zu jedem Bildmerkmal seine 5 Nachbarn nach dem Kriterium der Farbdistanz: $D(p, q) = \max(|r_p - r_q|, |g_p - g_q|, |b_p - b_q|)$
3. Interpoliere den Disparitätsraum zwischen je zwei benachbarten Bildmerkmalen.
4. Erweitere den Disparitätsraum nach unten und oben um einen fixen Wert: $[d_p, d_p + 1[\Rightarrow [d_p - \epsilon, d_p + 1 + \epsilon[$



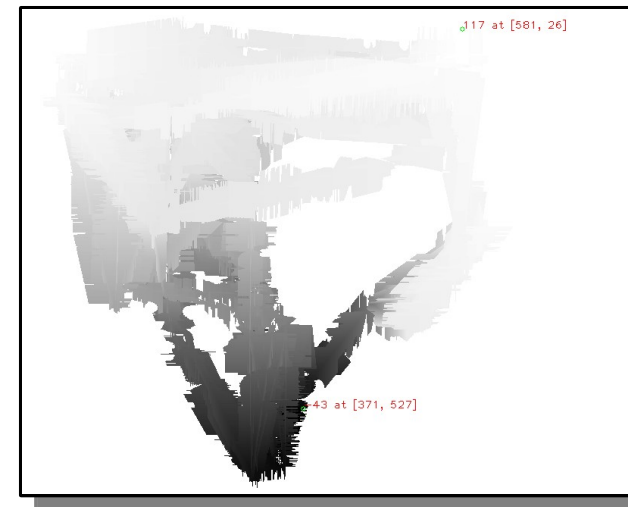
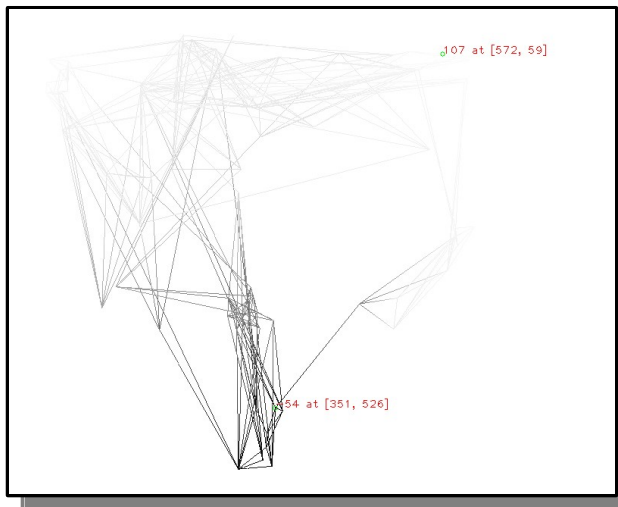
Abschätzung des Disparitätsraums



5. Verteile die gefundenen Disparitätsräume auf benachbarte Pixel derselben Unterstützungsregion.

- Unterstützungsregionen nach dem Modell der „Cross-based Aggregation“, d. h. in den vier Hauptrichtungen wird der Unterstützungsarm so lange verlängert, solange gilt:

1. $D_c(p_l, p) < \tau_1$ und $D_c(p_l, p_l + (1, 0)) < \tau_1$
2. $D_s(p_l, p) < L_1$
3. $D_c(p_l, p) < \tau_2$, wenn $L_2 < D_s(p_l, p) < L_1$



Abschätzung des Disparitätsraums

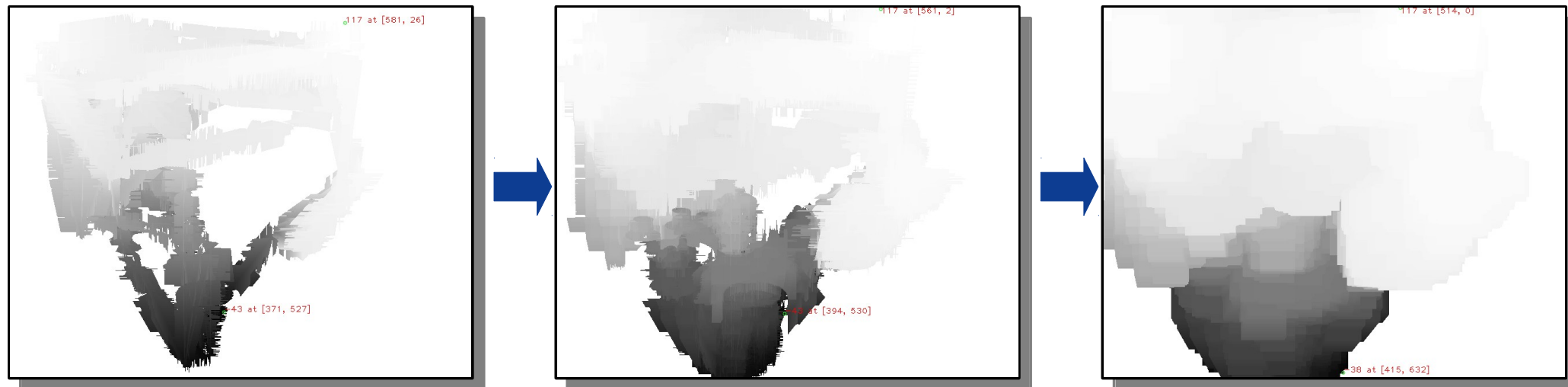


FernUniversität in Hagen
Lehrgebiet Mensch-Computer-Interaktion

6. Wiederhole diesen Schritt zwei Mal.

- ▶ [Mei2011]: Mei, Xing ; Sun, Xun ; Zhou, Mingcai et al.: On Building an Accurate Stereo Matching System on Graphics Hardware. In: GPU'11: ICCV Workshop on GPU in Computer Vision Applications, 2011. Online: <http://xing-mei.net/resource/pdf/adensus.pdf> (Abrufdatum: 2013-03-15).

7. Verteile den Disparitätsraum auf eine fixe Zahl von Nachbarpixeln in den vier Hauptrichtungen (Kohärenz).



Abschätzung des Disparitätsraums



FernUniversität in Hagen
Lehrgebiet Mensch-Computer-Interaktion

8. Führe den Stereo-Matching-Algorithmus für alle Pixel durch, für die bisher ein Disparitätsraum abgeschätzt wurde.
9. Verwende die errechneten Disparitäten zur Initialisierung des nächsten Iterationsschrittes, bis ein festgelegtes Abbruchkriterium (z. B. Berechnung von mind. 80 % der gültigen Pixel) eintritt.



Evaluation: Setting

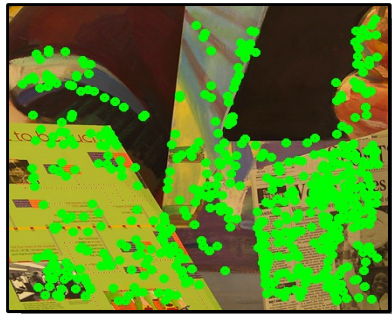


- Erprobt an 8 Bildpaaren:
 - 3 Bildpaare der auch in der Middlebury Stereo Evaluation verwendeten Middlebury Stereo Datasets:
 - Venus, Teddy, Cones
 - siehe <http://vision.middlebury.edu/stereo/eval/>
 - 5 Bildpaare, von Dr. Klaus Häming (Lehrgebiet Mensch-Computer-Interaktion) erstellt
 - 2x Off-Axis, 1x gekreuzt, 1x gekreuzt&gedreht, 1x hintereinander
 - Verwendung von überprüften (also korrekt zugeordneten) Paaren von Bildmerkmalen
 - Abbruchkriterium: Abschätzung des Disparitätsraums für mind. 80% der gültigen Bildpunkte erreicht, spätestens nach 5 Iterationen

Evaluation: Ergebnisse



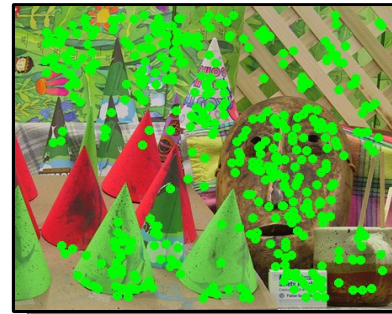
- Wenn viele Bildmerkmalspaare vorhanden: sehr gute Abschätzung
 - ▶ in 5 Fällen $\geq 80\%$ nach dem ersten Iterationsschritt
 - ▶ v. a. innerhalb der konvexen Hülle der Bildmerkmale



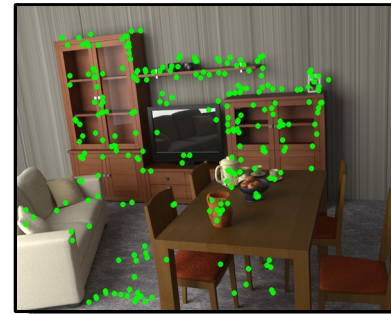
Venus



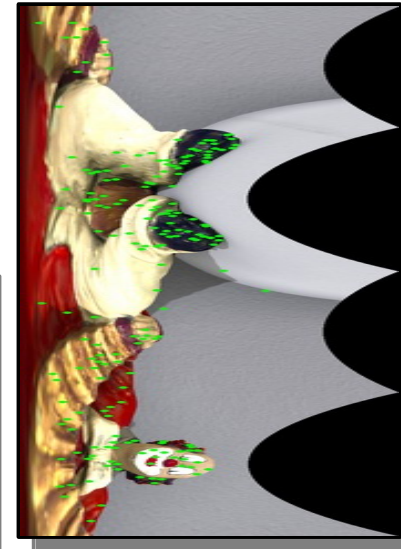
Teddy



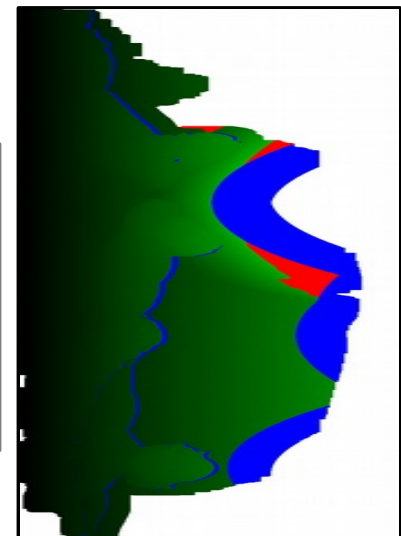
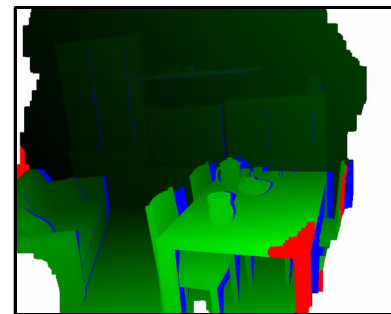
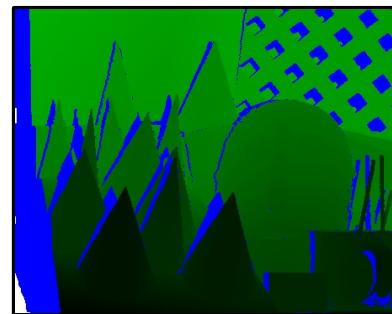
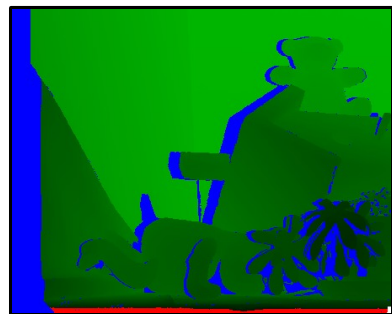
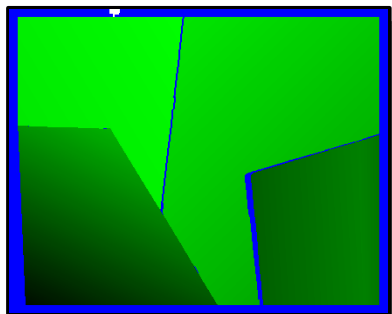
Cones



Raum Off-Axis



Clown-in-front

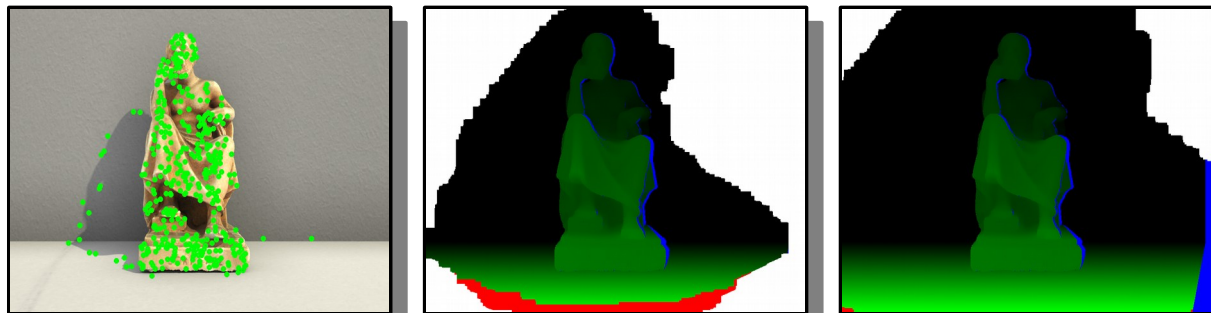


Evaluation: Ergebnisse

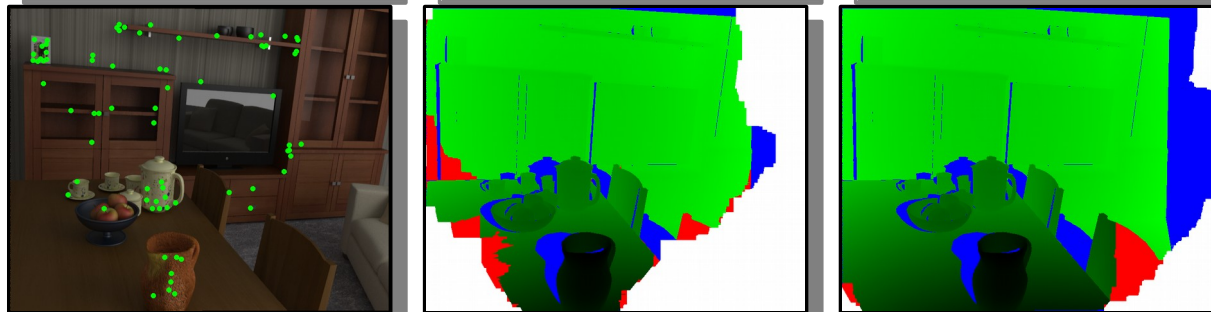


- In 2 Fällen 2 Iterationen, in einem Fall 4 Iterationen nötig
▶ Fehler werden tw. im nächsten Schritt korrigiert

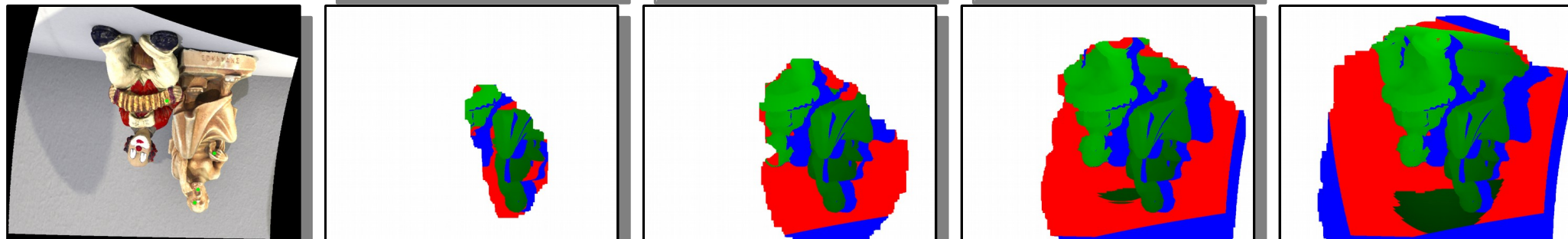
Sokrates



Wohnraum



Clown-Sokrates





- Algorithmus überträgt Disparitätsraum gut auf frontoparallele und wenig geneigte Flächen
- Algorithmus versagt bei schiefen Flächen und an Diskontinuitäten
 - ▶ Fehler an schiefen Flächen werden im nächsten Schritt gut korrigiert
 - ▶ An Diskontinuitäten versagt die Einschätzung oft auch im nächsten Schritt
- Schlechtestes Abschneiden beim Bildpaar „Clown-Sokrates“
 - ▶ nur 4 Bildmerkmalspaare
 - ▶ Grund: verdrehte Kameraachse
 - ▶ würde auch eine Rektifizierung unmöglich machen
- Berechnungsaufwand gegenüber Berechnung aller Disparitätsstufen von der minimalen bis zur maximalen (als bekannt vorausgesetzten) Disparitätsstufe
 - ▶ Venus 133,8 %; Teddy 53,8 %; Cones 57,3 %; Sokrates 91,2 %; Raum Off-Axis 63,2 %; Clown-Sokrates 17,5 %; Wohnraum 38,1 %; Clown-in-front 41,4 %
 - ▶ in fast allen Fällen starke Einsparung von Rechenaufwand
 - ▶ Ausnahme „Venus“: kleinstes Bildpaar mit wenigsten (20) Disparitätsstufen

- ▶ Wie könnte die Bestimmung benachbarter Bildmerkmale vereinfacht werden?
- ▶ Wie könnte die Abschätzung des Disparitätsraums an den Diskontinuitäten verbessert werden?
- ▶ Wie schneidet der Algorithmus im Vergleich zu einem Down-/Upscaling-Vorgehen ab?
- ▶ Wie im Vergleich zu anderen Algorithmen?
Vgl. z. B. Jan Cech: Growing Correspondence Seeds – A Fast Stereo Matching of Large Images. Online: <http://cmp.felk.cvut.cz/~cechj/GCS/> (Abrufdatum: 2013-03-17).

Bildnachweis, Sourcecode



FernUniversität in Hagen
Lehrgebiet Mensch-Computer-Interaktion

- „3D camera pistol grip“, © „3D shoot“ auf flickr (CC BY-NC-SA 2.0)
- Bilder „Venus“, „Teddy“, „Cones“:
 - Middlebury Stereo Datasets
 - vgl. <http://vision.middlebury.edu/stereo/data/>
- Bilder „Clown-in-front“, „Clown Sokrates“, „Sokrates“, „Raum Off-Axis“, „Wohraum“:
 - von Dr. Klaus Häming, Lehrgebiet Mensch-Computer-Interaktion, Fernuniversität in Hagen, für die Evaluationszwecke dieser Masterarbeit erstellt
 - mithilfe von 3D-Modellen und Texturen von Carlos Folch, <http://resources.blogscopia.com/>
 - und Texturen von Antti Lehtinen, <http://www.polygonblog.com/seamless-texture/>
- Quelltext: Doppellizenz (GPL und CC BY-NC-SA 3.0 Austria)
 - git@bitbucket.org:ChristophDrexler/masterarbeit.git