

High-Fidelity Image-Based Modeling

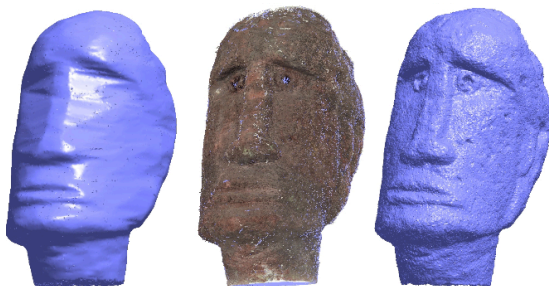
Steffen Heimes

Institut für Informatik
Humboldt-Universität zu Berlin

31. Mai 2011



Schritte zum Modell



1. Visuelle Hülle erstellen.
2. In Bildpaaren affin transformierte Features finden.
3. Diese bis zur maximalen Korrelation affin transformieren.
4. Ausgehend von den gefundenen Paaren (Patches) wachsen lassen.
5. Das Mesh anhand der gefunden Patches verbessern.



Gliederung

Übersicht

Erstellen der Eingangsdaten

- Visuelle Hülle
- Primäre Patches
- Sekundäre Patches

Verarbeitung

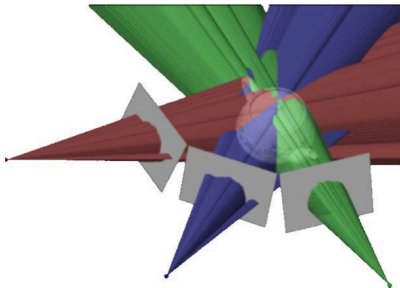
- Zwischenstand
- Verfeinerung des Meshes
- Multi-View Stereo Refinement

Fazit

- Implementierung und Resultate
- Quellen



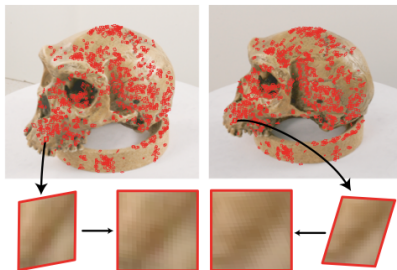
Visuelle Hülle



- ▶ Geschlossenes Dreiecksnetz.
- ▶ Kein Teilbereich der erstellte Oberfläche befindet sich innerhalb der zu rekonstruierenen Objektoberfläche.



Primäre Patches finden



- ▶ Durch die verschiedenen Betrachtungsrichtungen sind die Abbildungen lokal näherungsweise affin verzerrt.
- ▶ Mittels eines affin invarianten Detektor (Mikolajczyk & Schmid 2002) können auch diese Paare matchen.
- ▶ Affine Transformation, bis Kreuzkorrelation am größten ist.

Filtern der primären Patches

Falsch zueinander zugeordnete Paare werden hier entfernt.

- ▶ Durch Triangulieren wird die 3D-Position des Patches bestimmt und gleich verworfen, wenn die Position ausserhalb der visuellen Hülle ist.
- ▶ Bestimmung der geometrischen Konsistenz:
 - ▶ Zu jedem Match $P = (L, R)$ werden M nächste Nachbarn gesucht. $|M| = 5$
 - ▶ Aus der Menge der Nachbarn werden alle mögliche Tripel genommen und die baryzentrischen Koordinaten in dem von den Tripeln aufgespannten Dreieck bestimmt. (Jeweils in den Bildkoordinaten des jeweiligen Bildes)
 - ▶ Zwischen dem linkem und dem rechten Bild sollten die bestimmten Koordinaten nicht stark variieren.



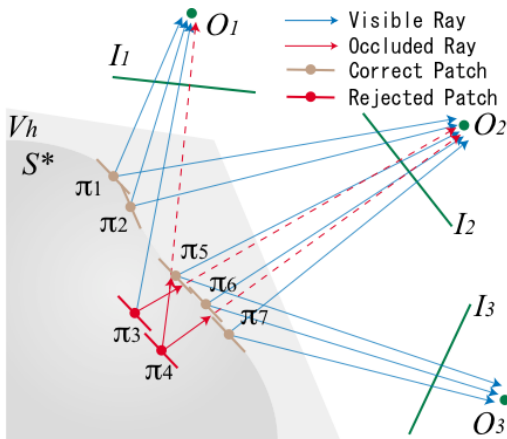
Expansion der Paare

- ▶ Die bislang gefundenen primären Patches sind nur lückenhaft auf der Oberfläche verteilt.
- ▶ Ausgehend von einem bekannten Patch $P = (L, R)$ mit einer Transformation \mathbf{T} wo gilt: $R = \mathbf{T}L$:
 - ▶ In der Umgebung von P wird ein Punkt L' genommen und anhand von \mathbf{T} zu R' transformiert.
 - ▶ Wenn die Korrelation zwischen L' und R' groß genug ist wird noch die Transformation \mathbf{T}' optimiert, bis Korrelation am größten ist.
 - ▶ Hinzufügen von $P' = (L', R')$ zur Menge von sekundären Patches.
- ▶ Wiederholen bis kein weiter Patch mehr hinzugefügt werden kann.



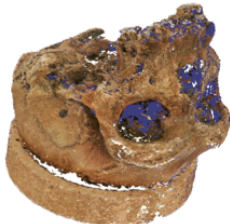
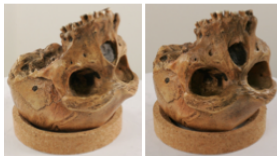
Verdeckte Patches filtern

- ▶ Test ob Paare innerhalb der wahrscheinlichen Oberfläche liegen.
- ▶ Paare werden aussortiert, wenn sie von Patches verdeckt werden, die näher an der Kamera sind.



Was haben wir?

- ▶ Die visuelle Hülle als erste Approximation, jedoch klar ausserhalb des zu erstellenden Modells.
- ▶ Eine Menge an Patches, wo die reale Oberfläche zu erwarten ist.



Wide-baseline Stereo Refinement

Minimierung der Energiefunktion $E = E_D + E_R + E_S$

E_D Abstand zur Oberfläche $E_D = \sum_{\vec{v}} D^2(\vec{v})/2$

E_R Silhouette consistency

E_S Smoothness

Dies wird erreicht indem jedes Vertex lokal über die Updateregeln $\delta\vec{v}$ den Gradienten entlang verschoben wird bis Konvergenz erreicht wird.

$$\vec{v} = \vec{v} + \delta\vec{v}$$

$$\begin{aligned} \delta\vec{v} = & -\max(-\beta_1, \min(\beta_1, \vec{D}(\vec{v})))\vec{N}(\vec{v}) \\ & + (\nabla E_R \cdot \vec{N}(\vec{v}))\vec{N}(\vec{v}) \\ & + (-\beta_2\Delta\vec{v} + \beta_3\Delta^2\vec{v}) \end{aligned}$$



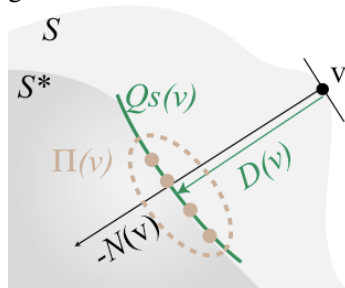
Abstand zur Oberfläche?

S Realexistierendes Mesh.

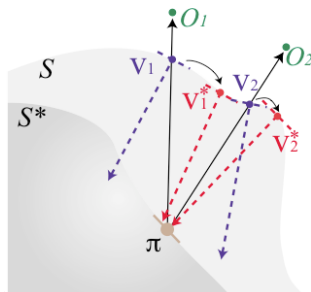
S^* Ideale Oberfläche.

$\hat{S}(\vec{v})$ Approximierte Oberfläche gebildet aus der Menge $\Pi(\vec{v})$.

$\Pi(\vec{v})$ ist die Menge an Patches, die für die Oberflächenapproximation von \vec{v} benutzt werden soll. Initialisiert mit \emptyset , wird sie inkrementell gefüllt.



Abstand zur Oberfläche?



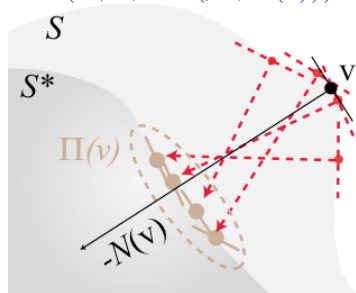
- ▶ Ausgehend von jedem Patch π , bestimmt durch zwei Fotos mit Zentrum O_1, O_2 , werden die nächsten Vertices \vec{v}_1, \vec{v}_2 vom Schnittpunkt des Strahls von π nach O_1, O_2 gewählt.
- ▶ In der Nachbarschaft von \vec{v}_i wird nun das Vertex \vec{v}_i^* mit dem kürzesten Weg von π zum von $\vec{N}(\vec{v}_i^*)$ definierten Strahl.
- ▶ π wird nun $\Pi(\vec{v}_i^*)$ hinzugefügt.



Abstand zur Oberfläche?

- ▶ Nun wird für jedes $\vec{v} \in S$ nun $\Pi(\vec{v})$ durch Hinzunahme der assoziierten Patches Nachbarverices von \vec{v} erweitert.
- ▶ Dies wird wiederholt, bis in $\Pi(\vec{v})$ 20 Patches enthalten sind oder nach 50 Iterationen.

$$\max(-\beta_1, \min(\beta_1, \vec{D}(\vec{v})))\vec{N}(\vec{v})$$



Kleine Details nachbilden

- ▶ Das Modell ist jetzt im groben hergestellt.
- ▶ Was fehlt sind die kleineren Details in der Oberfläche.
- ▶ Nur der erste Term wird verändert.

$$\vec{v} = \vec{v} + \delta\vec{v}$$

$$\begin{aligned} \delta\vec{v} = & -w(\vec{v})(\nabla f(\vec{v}) \cdot \vec{N}(\vec{v}))\vec{N}(\vec{v}) \\ & + (\nabla E_R \cdot \vec{N}(\vec{v}))\vec{N}(\vec{v}) \\ & + (-\beta_2\Delta\vec{v} + \beta_3\Delta^2\vec{v}) \end{aligned}$$

- ▶ $f(\vec{v})$ stellt hierbei die Abweichung der Oberfläche von dem dar, was in der Stereometrie gesehen wird.
- ▶ $w(\vec{v})$ ist ein Gewichtung, die bei jedem Vorzeichenwechsel von $f(\vec{v})$ verkleinert wird um Konvergenz zu erreichen.
- ▶ Konvergenz wird nach rund 20 Schritten erreicht.



Kleine Details nachbilden

- ▶ Das Modell ist jetzt im groben hergestellt.
- ▶ Was fehlt sind die kleineren Details in der Oberfläche.
- ▶ Nur der erste Term wird verändert.

$$\vec{v} = \vec{v} + \delta\vec{v}$$

$$\begin{aligned} \delta\vec{v} = & -w(\vec{v})(\nabla f(\vec{v}) \cdot \vec{N}(\vec{v}))\vec{N}(\vec{v}) \\ & + (\nabla E_R \cdot \vec{N}(\vec{v}))\vec{N}(\vec{v}) \\ & + (-\beta_2\Delta\vec{v} + \beta_3\Delta^2\vec{v}) \end{aligned}$$

- ▶ $f(\vec{v})$ stellt hierbei die Abweichung der Oberfläche von dem dar, was in der Stereometrie gesehen wird.
- ▶ $w(\vec{v})$ ist ein Gewichtung, die bei jedem Vorzeichenwechsel von $f(\vec{v})$ verkleinert wird um Konvergenz zu erreichen.
- ▶ Konvergenz wird nach rund 20 Schritten erreicht.



Kleine Details nachbilden

- ▶ Das Modell ist jetzt im groben hergestellt.
- ▶ Was fehlt sind die kleineren Details in der Oberfläche.
- ▶ Nur der erste Term wird verändert.

$$\vec{v} = \vec{v} + \delta\vec{v}$$

$$\begin{aligned} \delta\vec{v} = & -w(\vec{v})(\nabla f(\vec{v}) \cdot \vec{N}(\vec{v}))\vec{N}(\vec{v}) \\ & + (\nabla E_R \cdot \vec{N}(\vec{v}))\vec{N}(\vec{v}) \\ & + (-\beta_2\Delta\vec{v} + \beta_3\Delta^2\vec{v}) \end{aligned}$$

- ▶ $f(\vec{v})$ stellt hierbei die Abweichung der Oberfläche von dem dar, was in der Stereometrie gesehen wird.
- ▶ $w(\vec{v})$ ist ein Gewichtung, die bei jedem Vorzeichenwechsel von $f(\vec{v})$ verkleinert wird um Konvergenz zu erreichen.
- ▶ Konvergenz wird nach rund 20 Schritten erreicht.



Berechnung von $f(\vec{v})$

- ▶ Mittels Z-Buffering werden die Bilder bestimmt, wo \vec{v} sichtbar ist.
- ▶ Von diesen Bildern werden τ Bilder mit der geringsten perspektivischen Verzerrung gewählt. ($\tau = 5$)
- ▶ Über einen kleinen Ausschnitt auf Tangentialebene von S bei \vec{v} wird nun ein Grid mit $\mu \times \mu$ Punkten gelegt und von jedem der τ Bilder die Farbwerte von den Punkten bestimmt. ($\mu = 11$)
- ▶ Somit ergibt es τ Ausschnitte h_1, \dots, h_τ auf das Gebiet um \vec{v} .
- ▶ Die Farbwerte werden über die Durchschnittsintensität normalisiert.
- ▶ Nun wird die Summe der quadratischen Abweichungen über die Farbwerte berechnet.

$$f(\vec{v}) = \frac{2}{\tau(\tau-1)\mu^2} \sum_{i=1}^{\tau} \sum_{j=i+1}^{\tau} \text{SSD}(h_i, h_j)$$



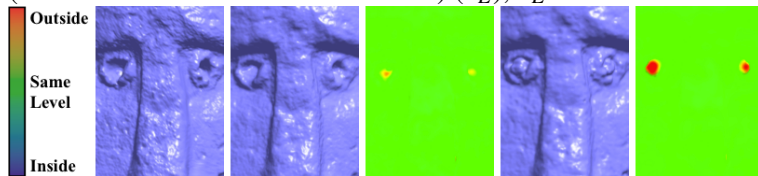
Implementierung

- ▶ Visuelle Hülle als gegeben vorausgesetzt. (Implementation nach (Lazebnik, Furukawa & Ponce 2007))
- ▶ Es wurden nur Bilder gematcht, deren Sichtachsen unterhalb eines bestimmten, angepassten Winkels lagen. ($10^\circ - 30^\circ$)
- ▶ Laufzeit der Berechnung auf einem 3.0 Ghz Pentium 4:
 - ▶ Die Matching Phase dauert 3 bis 20 Stunden.
 - ▶ Für die folgenden Verarbeitungsschritte werden 3 bis 4 Stunden benötigt.



Resultate

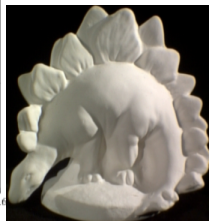
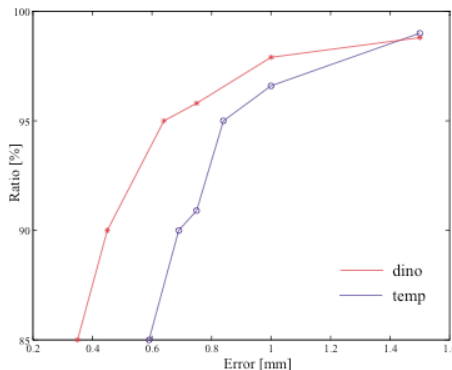
Bilder stellen von links nach rechts dar: Laserscan (S_L), Ergebnisse des hier vorgestellten Verfahrens (S), $S_L \rightarrow S$, Methode nach (Hernández Esteban & Schmitt 2004) (S_E), $S_E \rightarrow S$



- ▶ Geringe Differenz zum Laserscan.
- ▶ Bei den Augen ist auffällig, dass in S_E die Augen als konvex bestimmt wurden, real und hier richtig bestimmt sind sie konkav.



Resultate



- ▶ Vergleich zwischen S_L und S .
- ▶ Kumulatives Histogramm, auf x ist der maximale Fehler angeben. Der Wert in der y Achse gibt den prozentualen Anteil an Vertices an, deren Fehler kleiner ist.
- ▶ Keine Textur auf der Oberfläche des Dinos.







Ausblick

Die beschriebenen Patchfindungsalgorithmen und daraus resultierenden Oberflächenrekonstruktionen können auch in modifizierter Weise auf zufällig entstandene Bildserien angewendet werden.



Quellen

-  Furukawa, Y. & Ponce, J. (2006), High-fidelity image based modeling, Technical report, IN-RIA Rhone-Alpes.
URL: http://www.cs.washington.edu/homes/furukawa/papers/cvr_tr_2006_02.pdf
-  Hernández Esteban, C. & Schmitt, F. (2004), ‘Silhouette and stereo fusion for 3d object modeling’, *Computer Vision and Image Understanding* **96**(3), 367–392.
-  Lazebnik, S., Furukawa, Y. & Ponce, J. (2007), ‘Projective visual hulls’, *International Journal of Computer Vision* **74**(2), 137–165.
-  Mikolajczyk, K. & Schmid, C. (2002), ‘An affine invariant interest point detector’, *Computer Vision—ECCV 2002* pp. 128–142.

