

Gesteninterface

Ablauf

- **Allgemeines**
- Derzeitige Stand
- Probierte Ansätze & Weiteres Vorgehen

BV-Projekt 2013 / '14

Mirko Heger - Diplomstudent

Mirko.Heger@informatik.uni-hamburg.de

Was ich machen wollte ?

- Situation: Person sitzt vor dem Computer
- Die Hand im Bild finden
- Handgesten erkennen
 - Ausführen von Funktionen

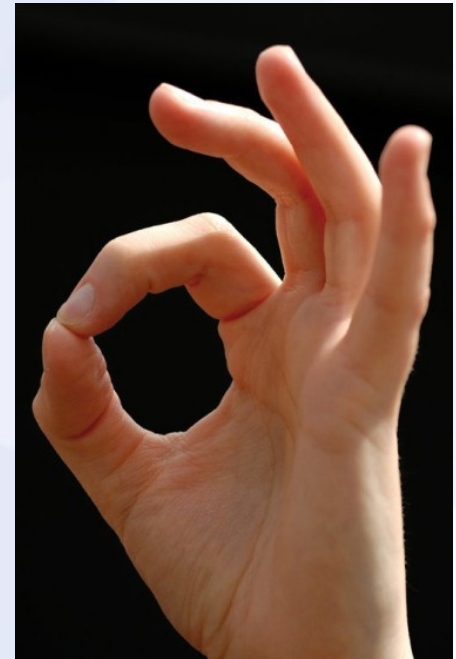
Was ich gemacht habe

- Einarbeitung in Numpy, Scipy, Opencv
- Kennen lernen der Kinect
- Ganzkörpergestenerkennung

Definition Gesten

- Lat. Gestus : Gebärdenspiel des Redners / Schauspielers
- Handlung mit symbolischem Charakter
- Bewegung, die etwas signalisiert
- Nonverbale Kommunikation

Gesten



Wie ich Gesten definiere

- Sichtbare Positionierung des Körpers
- Durch Kinect wahrnehmbare
 - Haltung
 - Positionierung
- Nicht zum Hintergrund gehörig

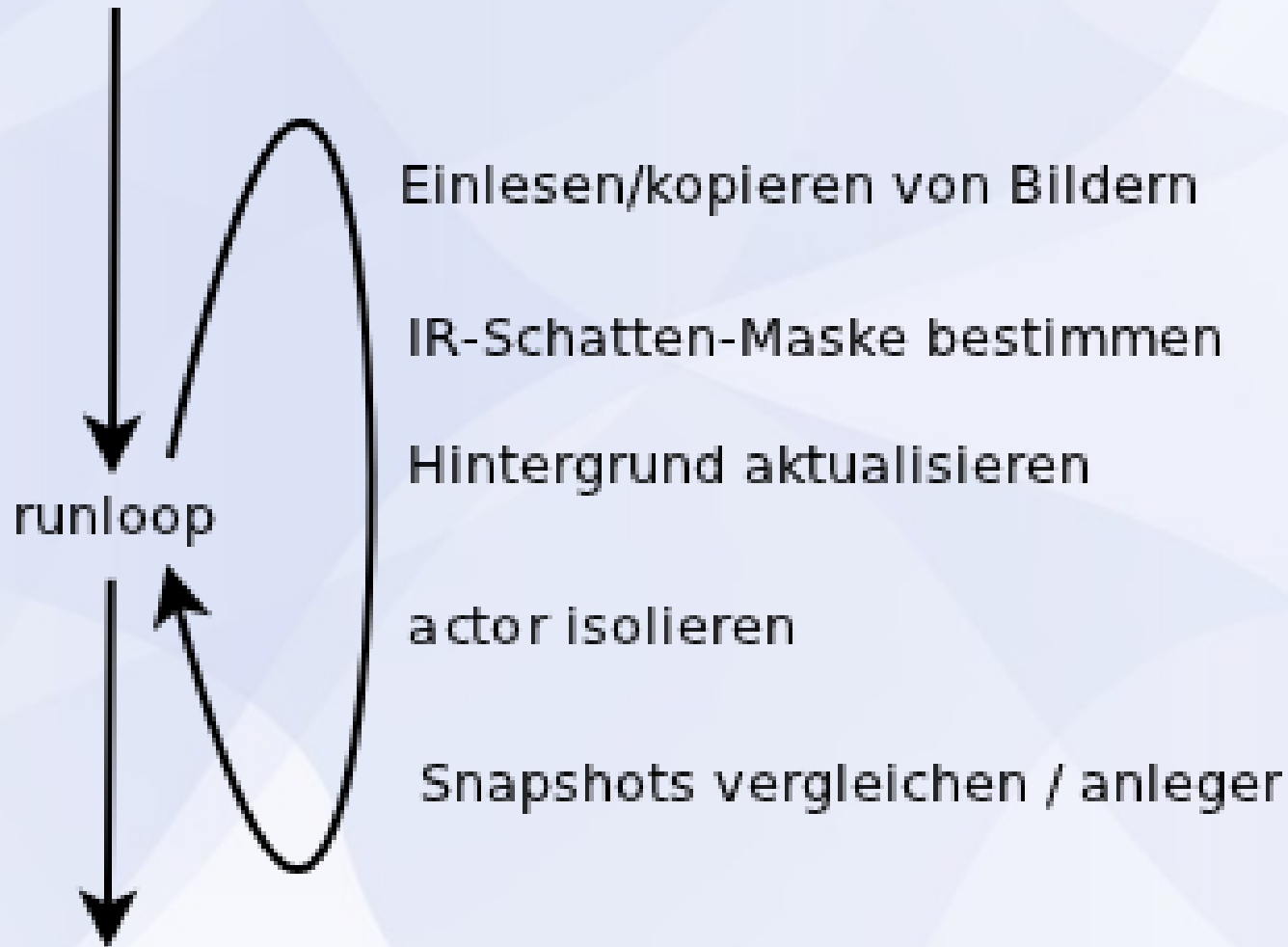
Meine Gesten



Ablauf

- Allgemeines
- **Derzeitige Stand**
- Probierte Ansätze & Weiteres Vorgehen

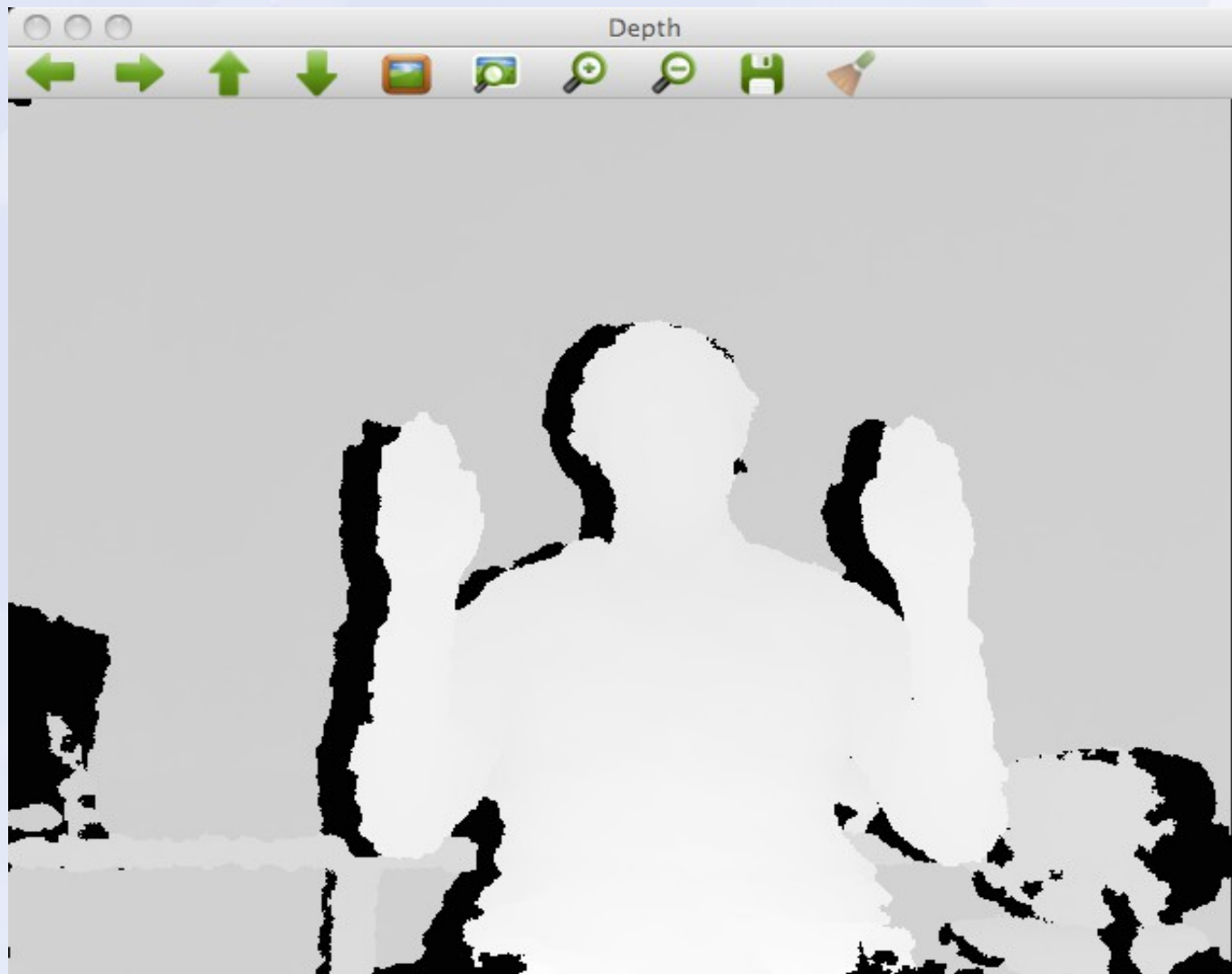
Programmaufbau



Einlesen und kopieren

- Bild einlesen und Kopie anlegen
 - Schutz vorm Überschreiben
- Datentyp anpassen

Tiefenbild

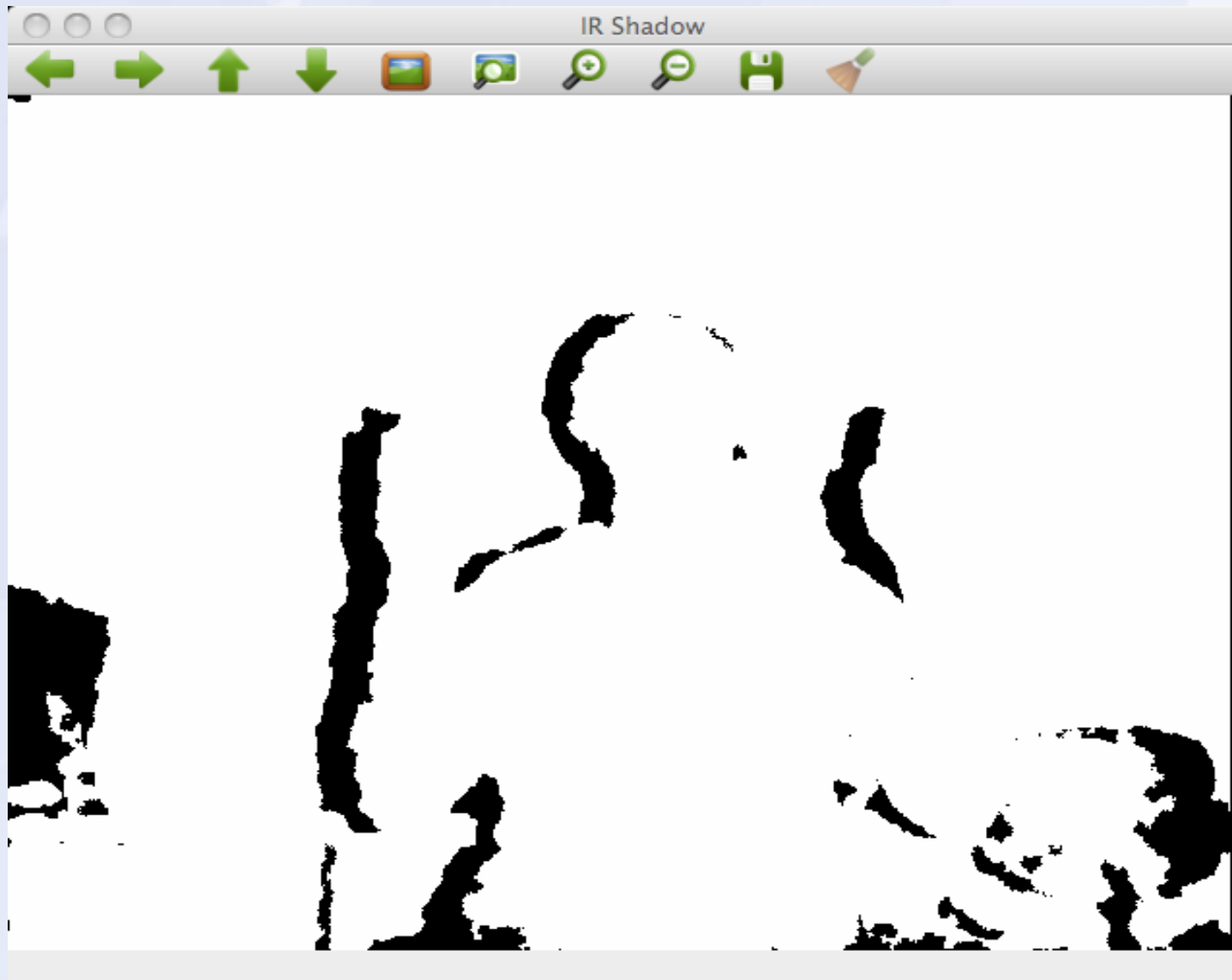


IR-Schatten-Maske bestimmen

- Annahme im Bild existiert IR-Schatten
- Größten Tiefenwert bestimmen
- Bit-Maske erstellen mit IR-Schatten = 0

- Schatten = 2047 ?
- Was wenn es ihn nicht gibt

IR-Schatten

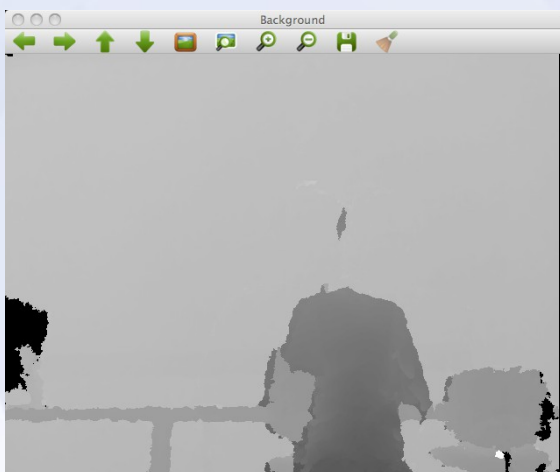
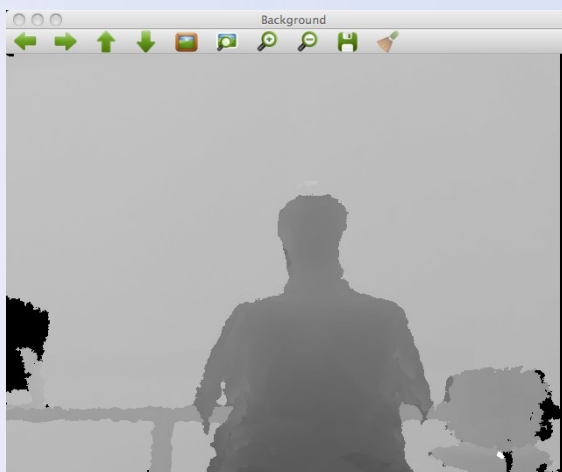
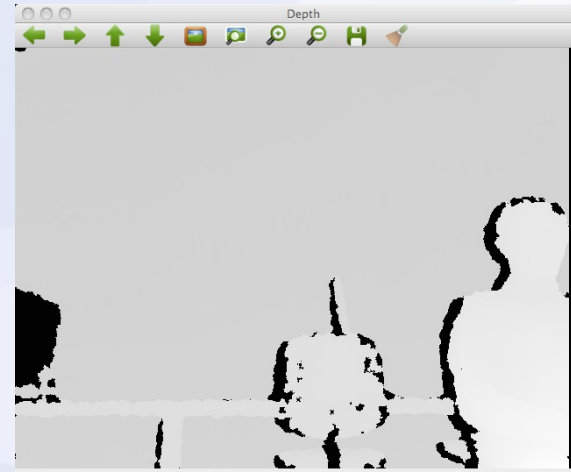
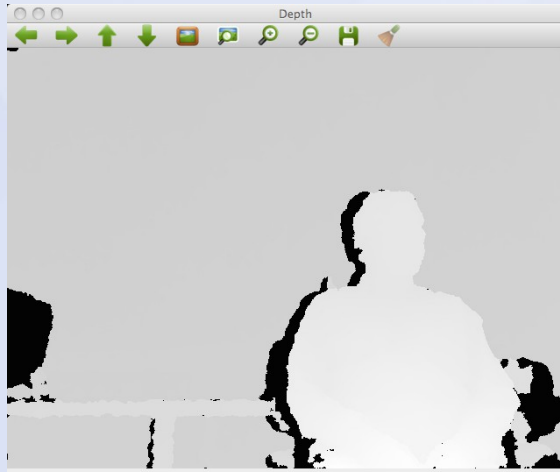
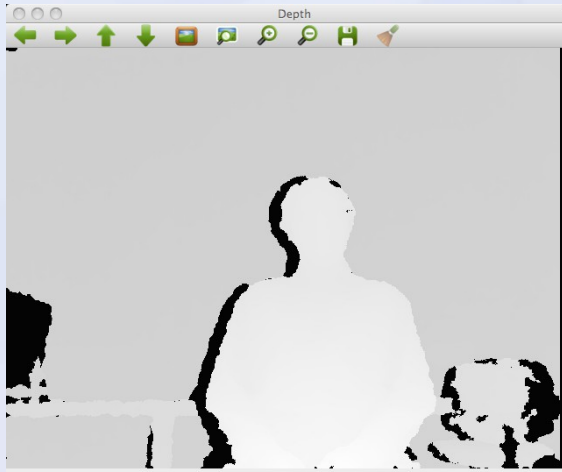


Hintergrund

- Hintergrund kumulativ aktualisieren
- Initialisierung:
Tiefe * IR-Maske
- Aktualisierung:
Tiefe * IR-Maske > Hintergrund → Neuer Wert

"Wenn ein Pixel, das nicht im IR-Schatten liegt, weiter entfernt ist als das bisher am weitesten entfernte Pixel an dieser Stelle, dann ist es eher die Hintergrundtiefe als das bisherige Pixel"

Kumulativer Hintergrund



Actor isolieren

Tiefenwert < Hintergrund -30:

Hintergrund

Sonst:

Actor

(Der Offset dient als Rauschschutz)

Weitere Filterung:

Histogramm bilden, 10 buckets,

Höchsten und tiefsten Bucket wegwerfen

Actor isolieren

Tiefenwert < Hintergrund -30:

Hintergrund

Sonst:

Actor

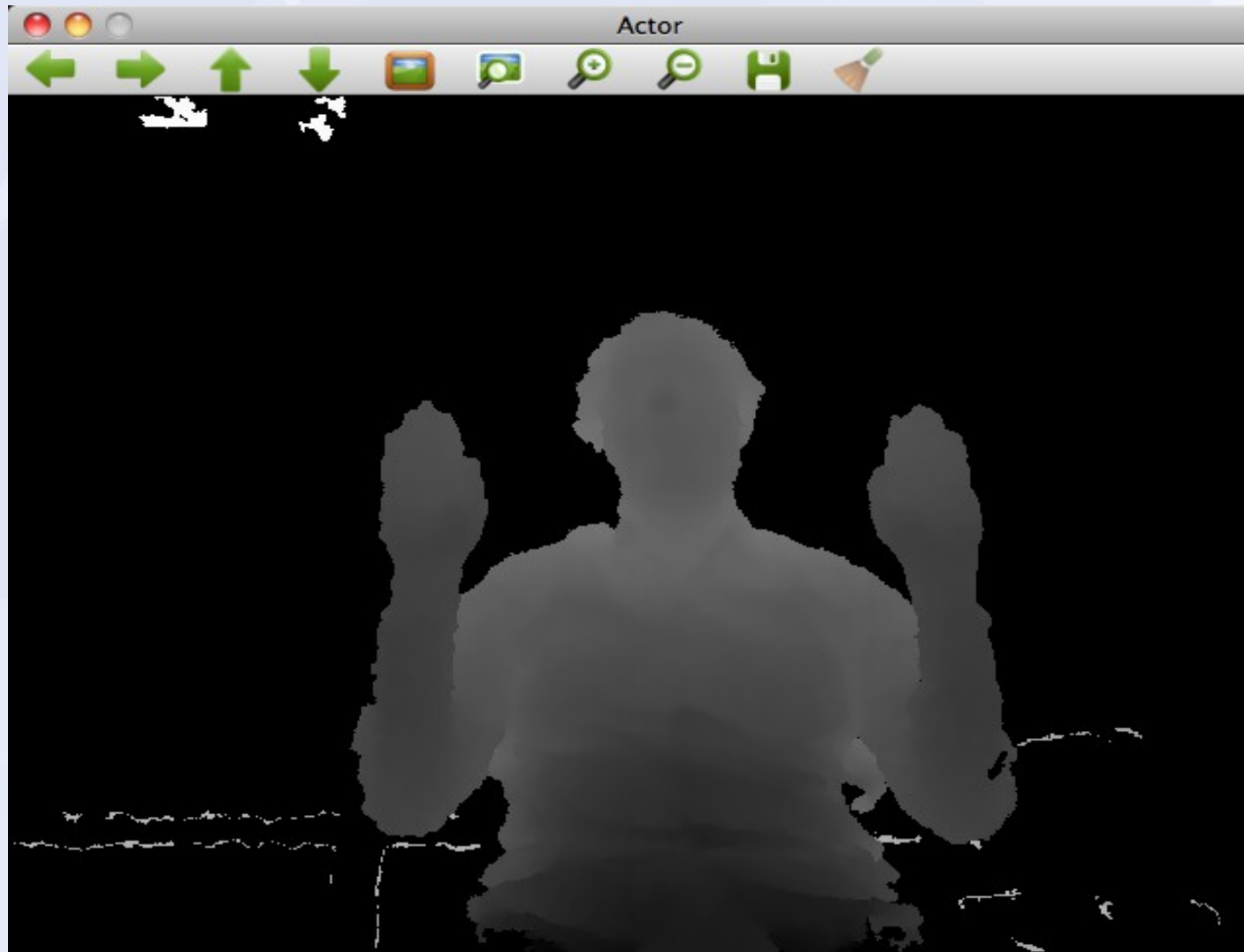
(Der Offset dient als Rauschschutz)

Weitere Filterung:

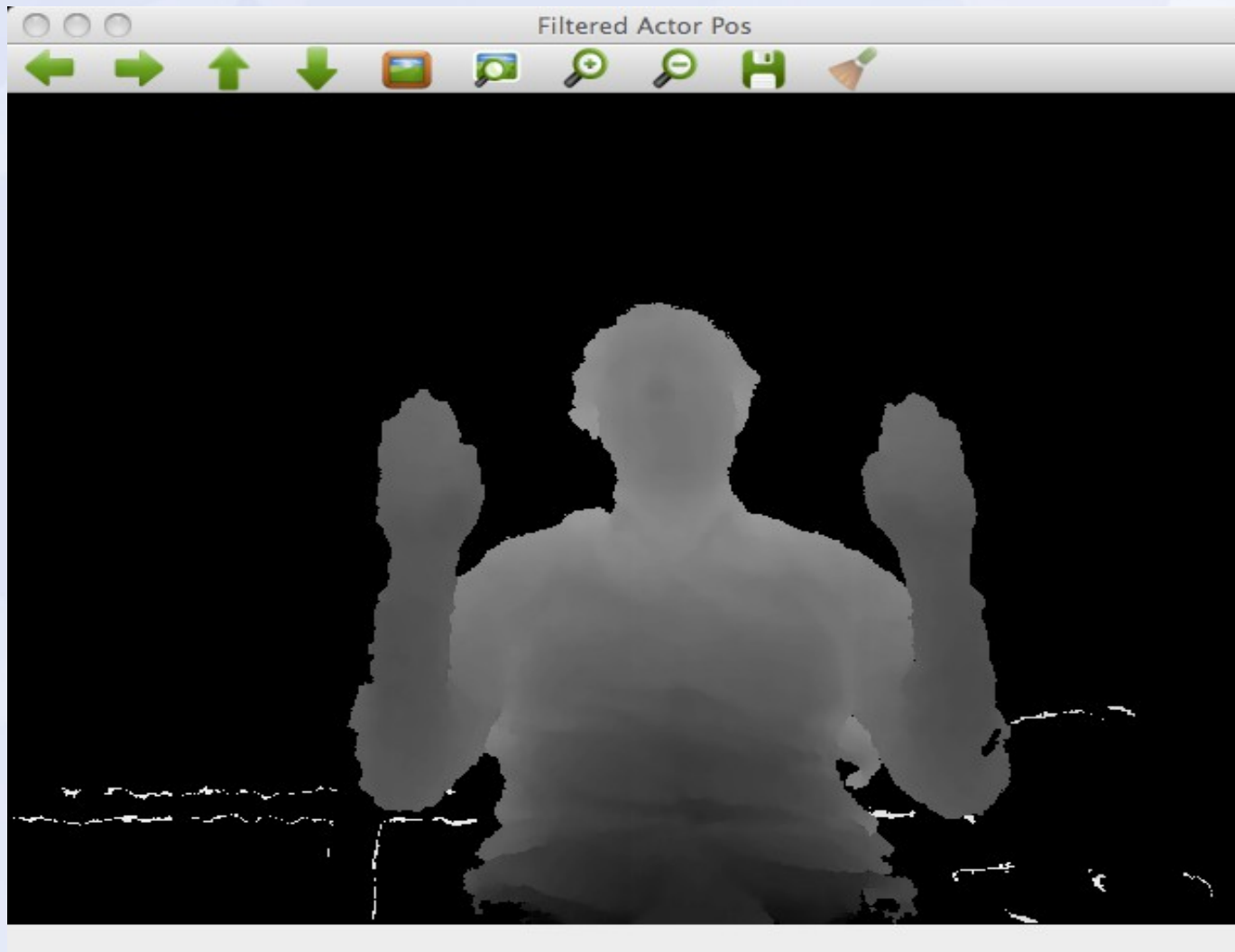
Histogramm bilden, 10 buckets,

Höchsten und tiefsten Bucket wegwerfen

Actor



Actor gefiltert



Snapshots anlegen / vergleichen



Hat der Benutzer "s" gedrückt, wird das aktuelle Bild als Gestenprototyp gemerkt

Es wird berechnet, welcher Gestenprototyp die größte Ähnlichkeit mit dem aktuellen actor-Bild hat

Für jeden Gestenprototyp wird

1. Die Differenz zum aktuellen Bild berechnet (und mit einem sehr kleinen Wert multipliziert, um int-Überläufe zu vermeiden)

2. Differenz wird quadriert, dann aufsummiert. Dieses ist das Differenzmaß

3. Ist das entstehende Differenzmaß kleiner als die bisherigen Differenzen, werden Differenz und Bild gemerkt

Ende der Schleife: Ähnlichstes Bild ist bestimmt

Snapshots anlegen / vergleichen



Hat der Benutzer "s" gedrückt, wird das aktuelle Bild als Gestenprototyp gemerkt

Es wird berechnet, welcher Gestenprototyp die größte Ähnlichkeit mit dem aktuellen actor-Bild hat

Für jeden Gestenprototyp wird

1. Die Differenz zum aktuellen Bild berechnet (und mit einem sehr kleinen Wert multipliziert, um int-Überläufe zu vermeiden)

2. Differenz wird quadriert, dann aufsummiert. Dieses ist das Differenzmaß

3. Ist das entstehende Differenzmaß kleiner als die bisherigen Differenzen, werden Differenz und Bild gemerkt

Ende der Schleife: Ähnlichstes Bild ist bestimmt

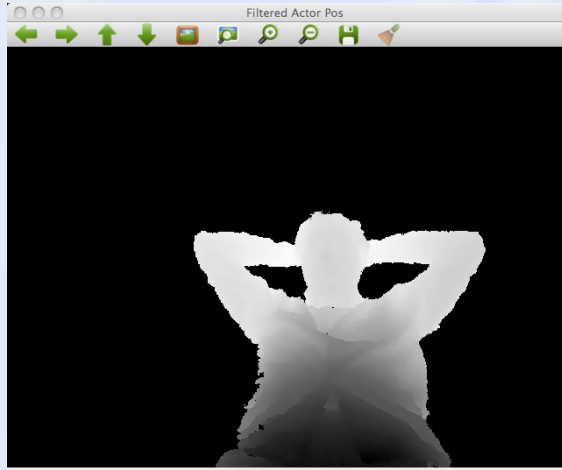
Snapshots anlegen

- Drückt der Benutzer „s“, wird ein Snapshot gespeichert
- Aufgespreizt über den Farbraum
- Als Paar gespeichert
 - Ein Negativ
 - Ein Positiv

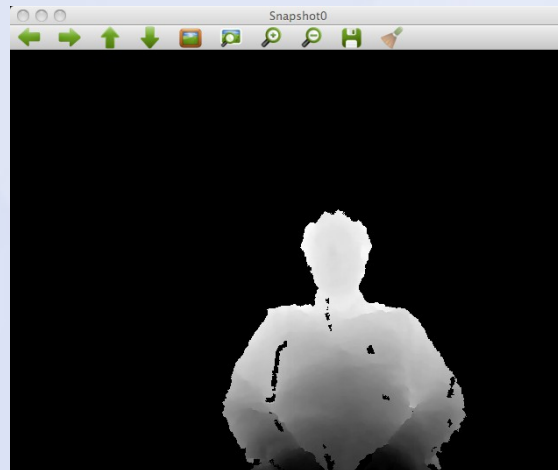
Snapshots vergleichen

- „Snapshot“ des Videostream erstellen
- Für jeden Snapshot:
 - Differenzbilder zum Stream erstellen
 - Differenzbilder quadrieren
 - Aufsummieren der einzelnen Bilder
 - Produkt der Summen
- Kleinste Differenz bestimmt den ähnlichsten Snapshot

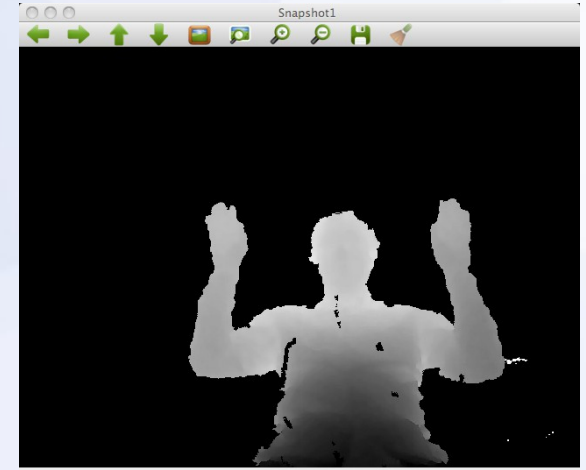
Snapshotts vergleichen



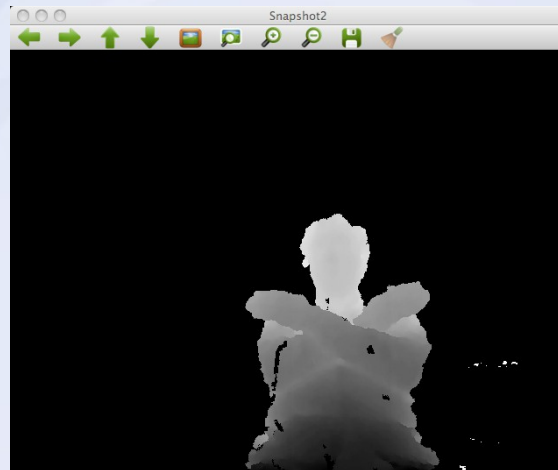
Videostream



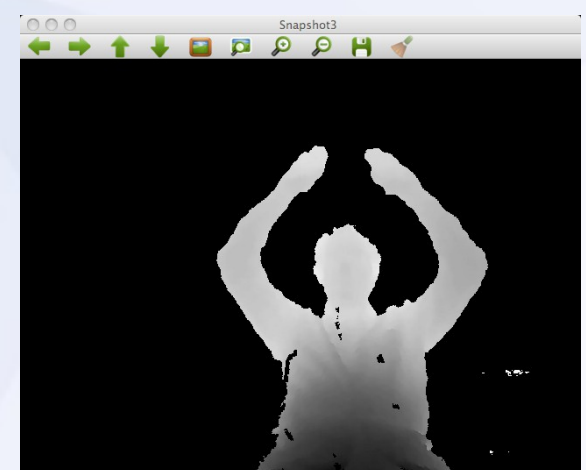
908,0928



1415,4807



750,2746



675,1847

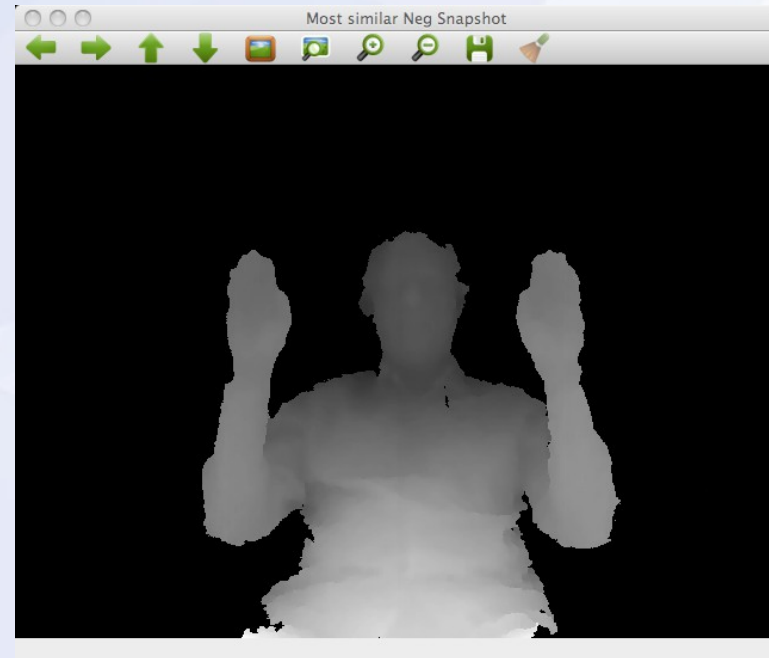
Was leistet die Wiedererkennung?

- Eine einfache Wiedererkennung, die unter Laborbedingungen gut funktioniert.

Snapshots

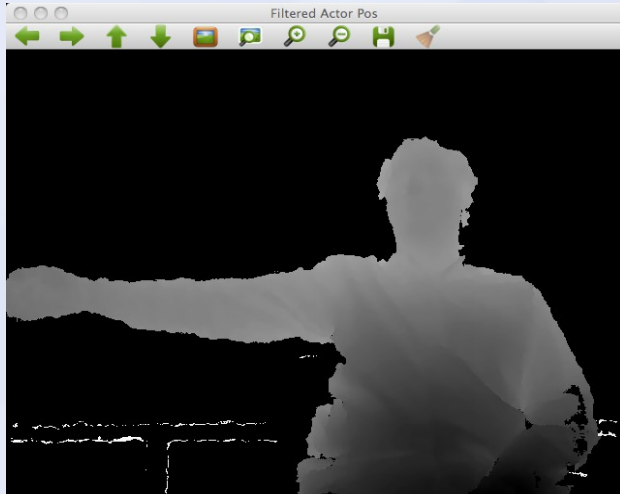


Videostream



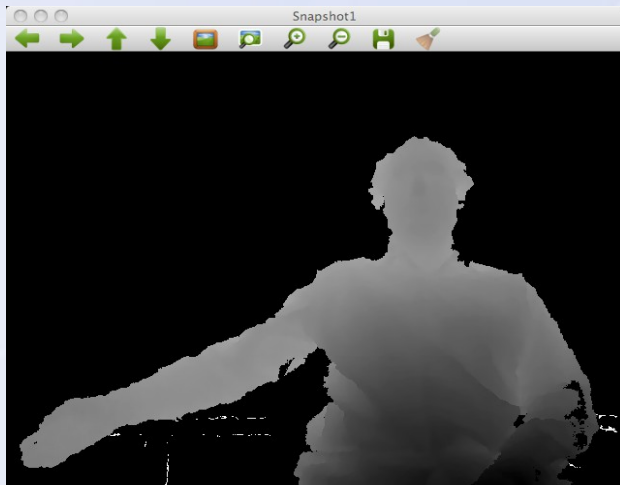
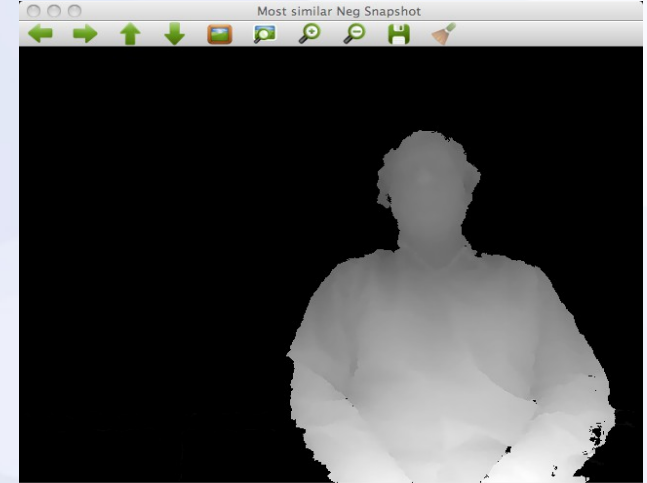
Wiedererkannte Geste

Nicht intuitive Wiedererkennung



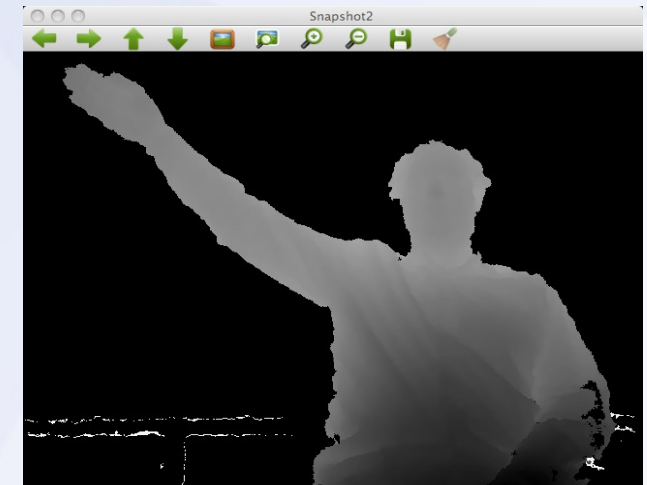
← Videostream

Erkannter
Snapshot →



← Snapshot 1

Snapshot 2 →



Was leistet der Algorithmus nicht ?

- Verschiebung des Actors
 - Hoch Runter
 - Seitwärts
- Vor und zurück
- Rotation wird nicht berücksichtigt

Was leistet das Programm, was nicht ?

- Echtzeitfähigkeit
- Ressourcen sparen

Ablauf

- Allgemeines
- Derzeitige Stand
- **Probierte Ansätze & Weiteres Vorgehen**

Ausprobierte Ansätze

- Boxed Actor – Box springt in der Form
- Trennung von Händen und Körper
 - Viel Körper im Bild, Handhaltung

Wie könnte man die Arbeit fortsetzen ?

- Bessere Rausch- und IR-Schattenunterdrückung
- Schwellwert für die Wiedererkennung
- Andere Methoden für die Wiedererkennung
- Boxed Actor
- Masseschwerpunkt → Bildverschiebung
- Folge von Bildern als Bewegung

Habt Ihr Vorschläge?