Automatic Hand Detection in RGB-Depth Data Sequences

Vitaly KONOVALOV, **Albert CLAPÉS**, *Sergio ESCALERA*

CCIA 2013

Universitat de Barcelona
HuPBA
CVC Computer Vision Center

Outline

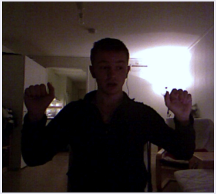
1. Introduction
2. Automatic Hand Detection method
   - Body segmentation
   - Reference point estimation
   - Body graph design
   - Geodesic distances map estimation
   - Including restrictions based on optical flow
   - Automatic hand detection
3. Results
   - Data, settings, and validation
   - Experiments
4. Conclusions

Motivation

Contextualization

- **Problem** Automatically detecting hands in RGB-D data (involving changes in illumination, viewpoint variations, and hand and wrist's articulated and deformable nature).

- **Scenario** People seated at the desk (only upper body is visible).



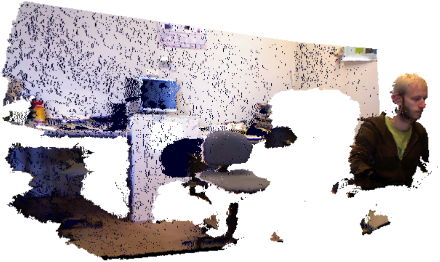
- **Assumptions** (a) Upper body frontal view, (b) non-cluttered desk, and (c) hand landmarks remain at constant geodesic distance from an anatomical reference point.

Applications

- Natural human-computer interaction, gaming, and monitorization.

State-of-the-art

- Two paradigms for **body part detection in visual data**:
  - Learning based approaches.
  - Body parameter estimation from observed features.
- Typically, computer vision methods were relying on RGB information.
- **Hardware devices** for depth (or rgb-depth) data acquisition:
  - Structured Light (Microsoft® Kinect™).
  - Time-of-Flight.



Outline

1 Introduction

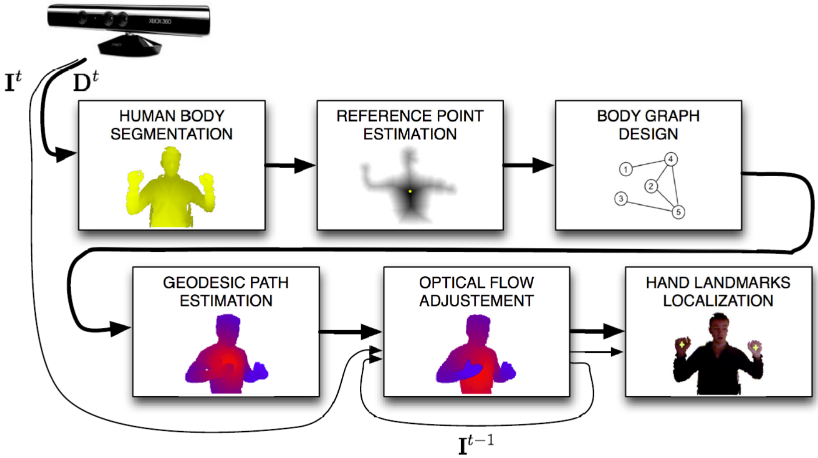
2 Automatic Hand Detection method

- Body segmentation
- Reference point estimation
- Body graph design
- Geodesic distances map estimation
- Including restrictions based on optical flow
- Automatic hand detection

3 Results

4 Conclusions

Automatic Hand Detection method



Outline

1 Introduction

2 Automatic Hand Detection method
- Body segmentation
- Reference point estimation
- Body graph design
- Geodesic distances map estimation
- Including restrictions based on optical flow
- Automatic hand detection

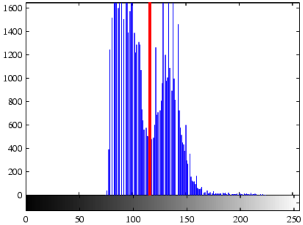
3 Results
- Data, settings, and validation
- Experiments

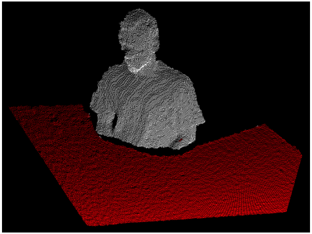
4 Conclusions

Body segmentation

Given a new depth frame $\mathbf{D}^t$, in which a subject seated in a non-cluttered desk appears,

1. **Foreground segmentation**
   Otsu's thresholding on depth values, assuming a bimodal distribution (foreground vs background).
2. **Point cloud representation**
   Project the foreground in a cloud of 3D points $\mathbf{P}^t$.
3. **Table extraction**
   The table plane is modeled using *RANSAC*, featuring the points in $\mathbf{P}^t$ by their surface normal vectors.





Outline

1 Introduction

2 Automatic Hand Detection method
- Body segmentation
- Reference point estimation
- Body graph design
- Geodesic distances map estimation
- Including restrictions based on optical flow
- Automatic hand detection

3 Results
- Data, settings, and validation
- Experiments

4 Conclusions

Reference anatomical landmark (torso point)

Given the point cloud $\mathbf{B}^t$, representing the segmented human body segmented from $\mathbf{P}^t$,

1. Re-project $\mathbf{B}^t$ into a 2D dense depth image and compute a **distance map** within the body region.
2. Compute the reference point $\mathbf{x}^t_{\mathrm{ref}}$. Let C be the set of contour points in the silhouette, then

$$\mathbf{x}^t_{\mathrm{ref}} = \mathrm{argmax}_{\mathbf{x}} \mathrm{min}_{\mathbf{x}_C \in C} d(\mathbf{x}, \mathbf{x}_C)$$

Outline

1 Introduction

2 Automatic Hand Detection method
- Body segmentation
- Reference point estimation
- Body graph design
- Geodesic distances map estimation
- Including restrictions based on optical flow
- Automatic hand detection

3 Results
- Data, settings, and validation
- Experiments

4 Conclusions

Body graph design

1. Project the 2D point reference point $\mathbf{x}^t_{\mathrm{ref}}$ to a 3D point $\mathbf{p}^t_{\mathrm{ref}}$ (from which the graph construction will start).
2. Partitionate $\mathbf{B}^t$ in a voxel grid.
3. The body graph $G^t = (V^t, E^t, W^t)$ is constructed as follows:

$$V^t = \{\mathbf{q}_{ijk} : \mathbf{q} \text{ is the centroid of the points } \mathbf{p} \text{ of } \mathbf{B}^t \text{ in the } (i,j,k) \text{ voxel}\}$$

$$E^t = \{(\mathbf{q}_{ijk}, \mathbf{q}_{i'j'k'}) \in V^t \times V^t : \| (i,j,k)^T - (i',j',k')^T \|_\infty < 1\}$$

$$W^t = \{w(e) = ||\mathbf{q} - \mathbf{q}'||_2 : e = (\mathbf{q}, \mathbf{q}') \in E^t\}$$

that is, two points $\mathbf{q}$ and $\mathbf{q}'$ are connected by an edge if they are in the same 3D neighborhood of $3 \times 3 \times 3$ voxels.

Outline

1. Introduction
2. Automatic Hand Detection method
   - Body segmentation
   - Reference point estimation
   - Body graph design
   - Geodesic distances map estimation
   - Including restrictions based on optical flow
   - Automatic hand detection
3. Results
   - Data, settings, and validation
   - Experiments
4. Conclusions

Geodesic distances map estimation

- Using G^t, it is possible to measure geodesic distances between two different body locations.
- The **geodesic distance between two body locations** is the length of the shortest path over the body surface.
- The geodesic distance $d_G(\mathbf{q}, \mathbf{q}')$ is estimated:

$$d_G(\mathbf{q}, \mathbf{q}') = \sum_{e \in E_{SP}(\mathbf{q}, \mathbf{q}')} w(e)$$

where $E_{SP}(\mathbf{q}, \mathbf{q}')$ contains all the edges along the shortest path between $\mathbf{q}$ and $\mathbf{q}'$, computed using the min-path Dijkstra's algorithm.

Outline

1 Introduction

2 Automatic Hand Detection method
- Body segmentation
- Reference point estimation
- Body graph design
- Geodesic distances map estimation
- Including restrictions based on optical flow
- Automatic hand detection

3 Results
- Data, settings, and validation
- Experiments

4 Conclusions

Geodesic paths desambiguation (I)

Problem

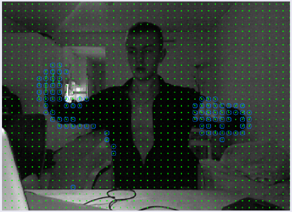
The arms are stick together or to another body part ⇒ undesired graph connections ⇒ bad geodesic paths estimations

Geodesic paths desambiguation (II)

Solution

Use the optical flow magnitude to remove undesired graph edges.

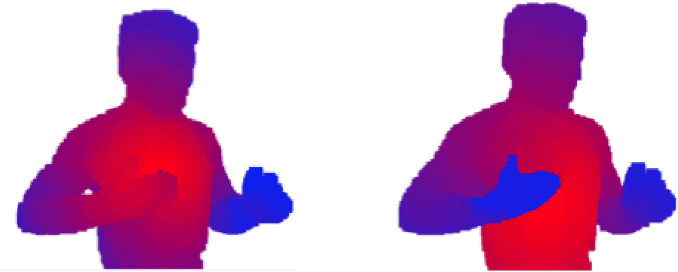
- At each time step, a dense optical flow map $\mathcal{F}^t$ is computed using I^t and I^{t-1}.



- Update E^t:

$$E^t := E^t - \{(\mathbf{x}_{ij}, \mathbf{x}_{kl}) \in E^t : \| \, |\mathcal{F}^t(i,j)| - |\mathcal{F}^t(k,l)| \, \|_2 > \beta\}$$

Geodesic paths desambiguation (III)



Outline

1 Introduction

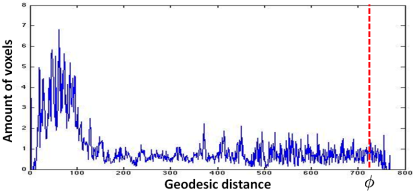
2 Automatic Hand Detection method
- Body segmentation
- Reference point estimation
- Body graph design
- Geodesic distances map estimation
- Including restrictions based on optical flow
- Automatic hand detection

3 Results
- Data, settings, and validation
- Experiments

4 Conclusions

Automatic hand detection

Once the geodesic map has been computed from G^t (considering optical flow restrictions), a **histogram of geodesic values** H_{G^t} is computed.



The **hands' regions are segmented** as those two biggest connected components in which their points have a value greater than ϕ. Being $\phi = g(H_{G^t}, \theta)$, and θ an experimental parameter expressing the 'percentage' of highest geodesic distance values.

The **hands are located** at the 3D centroid of the segmented regions.

Outline

1 Introduction

2 Automatic Hand Detection method

3 Results
- Data, settings, and validation
- Experiments

4 Conclusions

Outline

1 Introduction

2 Automatic Hand Detection method
- Body segmentation
- Reference point estimation
- Body graph design
- Geodesic distances map estimation
- Including restrictions based on optical flow
- Automatic hand detection

3 Results
- Data, settings, and validation
- Experiments

4 Conclusions

Data, settings, and validation

- **Data**
  - RGB-D dataset recorded with a Kinect™.
  - 6 different users simulating upper-body HCI scenarios.
  - 3000 RGB-D frames at 640×480 resolution.
  - Groundtruth: 2171 manually annotated hands (interactive 3D viewer).
- **Settings** $s = 20\,\mathrm{mm}$, $\beta = 0.2$, $\gamma = 25\,\mathrm{mm}$, and $\theta = 1\%$
- **Validation** Detection accuracy based on groundtruth hands and different tolerance values for β, γ, and θ. Posterior fine-tuning of γ and θ.

Outline

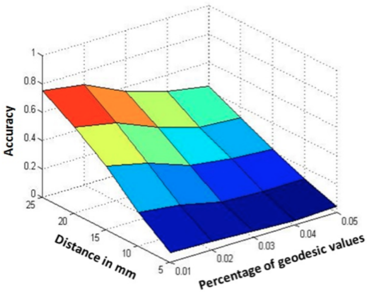
1 Introduction

2 Automatic Hand Detection method
- Body segmentation
- Reference point estimation
- Body graph design
- Geodesic distances map estimation
- Including restrictions based on optical flow
- Automatic hand detection

3 Results
- Data, settings, and validation
- Experiments

4 Conclusions

Quantitative results: detection accuracy



	Geodesic path	Geodesic path (w/ optical flow)
Accuracy	74.12%	**84.15%**

Qualitative results (I): geodesic distance map estimations

Figure : Color images, depth maps, estimated geodesic maps, and labeled hands.

Qualitative results (II): a sequence of detections

Figure : Detection in a sequence of frames.

Outline

1. Introduction
2. Automatic Hand Detection method
3. Results
4. Conclusions

Conclusions

- Simple and fully-automatic.
- Robust to partial occlusions (detection in still images, not relying in tracking strategies).
- Not requiring large training datasets and training phase.
- More cumbersome body segmentation strategies to deal with clutter in desks.
- Thinking in possible improvements for efficency.
- FYI

PrimeSense OpenNI® pcl

Thank you for you attention!