Robust Model-based 3D Head Pose Estimation

Gregory P. Meyer[1,2] Shalini Gupta[2] Iuri Frosio[2] Dikpal Reddy[2] Jan Kautz[2]

gmeyer3@illinois.edu shalinig@nvidia.com ifrosio@nvidia.com dikpal@gmail.com jkautz@nvidia.com

[1]University of Illinois at Urbana-Champaign [2]NVIDIA

Motivation

Estimating the 3D pose of a head has many important **applications**:

- Facial motion capture
- Human-computer interaction
- Video conferencing

In addition, it is **a pre-requisite to** several problems:

- Remote Gaze tracking
- Face recognition
- Face expression analysis

Head pose estimation has **traditionally been performed on RGB** images by detecting facial features. However, this can be **difficult when illumination variations, shadows, and occlusions** are present. With the emergence of **inexpensive commodity depth cameras**, promising 3D techniques for body, hand, and head pose estimation have been proposed.

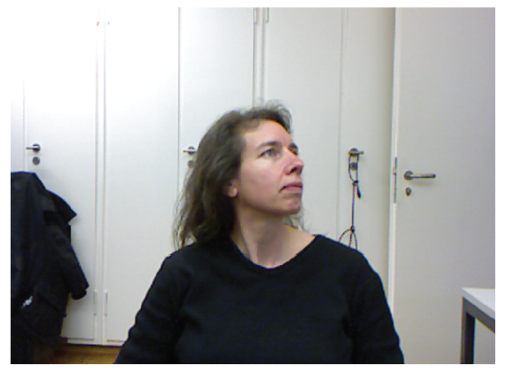
Contributions

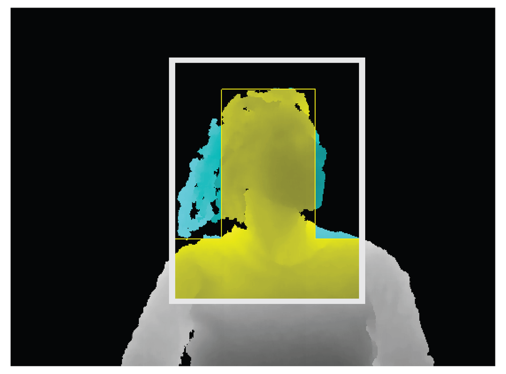
We developed a method for accurate 3D head pose estimation for data acquired with commodity depth cameras. Our approach uses only 3D information, requires no manual intervention or training, and generalizes well to different 3D sensors. We achieve state-of-the-art results on the Biwi Kinect dataset by combining a number of concepts:

- We detect the head using an adaptive 3D matched filter.
- We estimate the pose by registering a morphable face model to the measured facial data through a combination of particle swarm optimization (PSO) and the iterative closest point (ICP) algorithm.
- Also, we continuously adapt the morphable model to fit the subject's face, and dynamically weight the model to give more importance to the useful visible parts of the face.

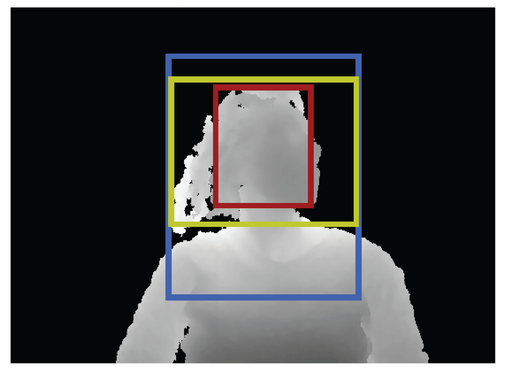
Head Localization

- Head localization identifies an area within a depth image which most likely contains a head.
- We use an adaptive detection filter, whose size changes to match the expected size of an average human head at various depths.


Input Image


Adaptive Filter


Score Image


Detected Head Location

Head Pose Estimation

We estimate the head pose by identifying the rotation and translation of the morphable model that best matches the observed data. We accomplish this by minimizing the following cost function:

$$E(\mathbf{x}) = E_v(\mathbf{x}) + \lambda E_c(\mathbf{x})$$

- $\mathbf{x} = (\theta_x, \theta_y, \theta_z, t_x, t_y, t_z)$ is the pose of the model.
- $E_v(\mathbf{x})$ measures the distance between corresponding vertices in the model and observed depth image.
- $E_c(\mathbf{x})$ penalizes the pose when there are too few corresponding points.

We update the shape and weights of the morphable model to match the observed data:

- Deform the surface of the morphable model to minimize the difference between the model vertices and observed data.
- Weight the vertices of the model proportional to the residual difference.

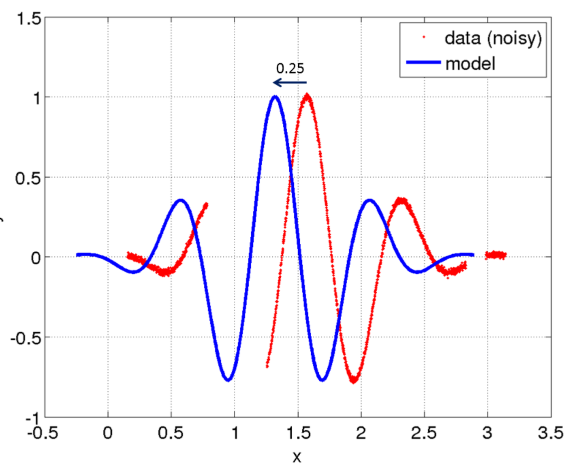
PSO and ICP Optimization

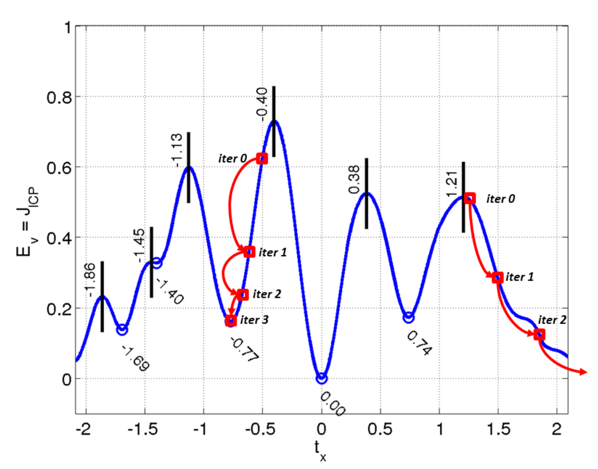
Our approach uses a combination of particle swarm optimization (PSO) and the iterative closest point (ICP) algorithm to minimize the cost function $E(\mathbf{x})$:

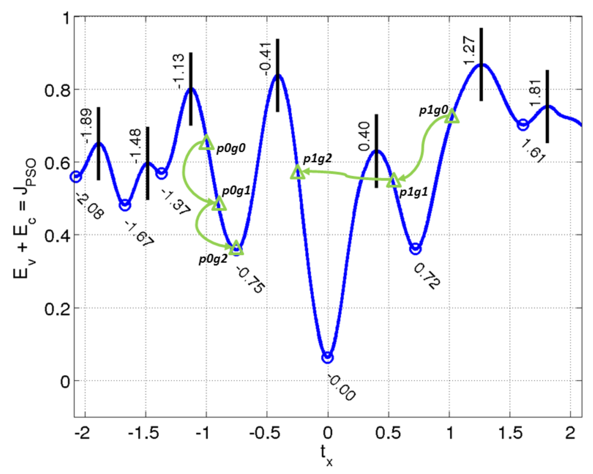
- The particles stochastically sample the cost function searching for the global optimum.
- ICP accelerates the search by pushing the particles to local optimum.

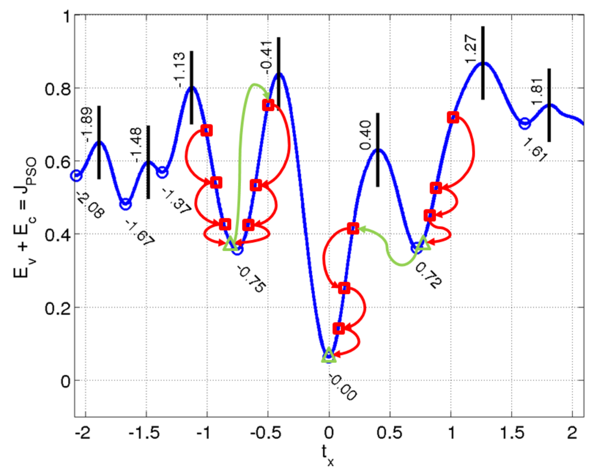
A 1D example to visualize the interaction between PSO and ICP:

- ICP quickly converges to a local minima, but it is sensitive to its starting position.
- PSO searches a large range of parameters, but is prone to premature convergence.
- Our method combines the two techniques to overcome their individual shortcomings. PSO enables our method to search a large range of possible poses, and ICP constrains our search to local minima in the parameter space.


Registration Problem


ICP


PSO


PSO + ICP

Results

We measured the performance of our method and compared it with state-of-the-art algorithms on the Biwi Kinect Head Pose dataset. The dataset was acquired with a Kinect sensor and contains over 15K RGB and depth images of 20 subjects. Our proposed algorithm produced the lowest rotational errors compared to existing methods.

Method	Errors			
	Yaw [°]	Pitch [°]	Roll [°]	Location [mm]
Proposed (ICP Only)	3.7	4.3	4.0	11.1
Proposed (PSO Only)	7.7	8.4	6.0	23.3
Proposed (PSO+ICP)	2.1	2.1	2.4	5.9
Fanelli	8.9	8.5	7.9	14.0
Tulyakov	4.7	7.6	5.3	9.1
Padeleris	11.1	6.6	6.7	13.8
Martin	3.6	2.5	2.6	5.8
Rekik	5.1	4.3	5.2	5.1
Baltrusaitis	6.3	5.1	11.3	7.6
Papazov	3.9	3.0	2.5	8.4