Head Pose Estimation in Computer Vision: A Survey

Erik Murphy-Chutorian, Student Member, IEEE,
and
Mohan Manubhai Trivedi, Fellow, IEEE
PAMI, April 2009

Presentation by: Naresh Kumar M S

Introduction

- What is Head Pose Estimation?
- Challenges – camera distortion, illumination changes, facial expressions, biological appearance, occlusion, perspective distortion.

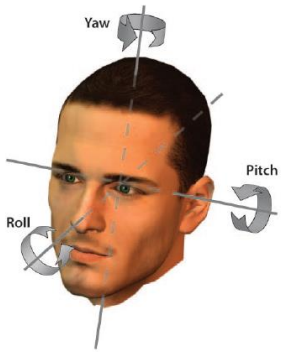

Fig. 1. The three degrees of freedom of a human head can be described by the egocentric rotation angles *pitch*, *roll*, and *yaw*.

Motivation

- Head movement is an important gesture.
- Useful to infer other information such as gaze.
- Applications in HCI, Driver Assistance Systems, Activity and Behavior Analysis.

(a) (b)

Fig. 2. Wollaston illusion: Although the eyes are the same in both images, the perceived gaze direction is dictated by the orientation of the head [134].

Head Pose Estimation - Methods

- 8 categories based on the fundamental approach used:
  -Appearance Template methods
  -Detector Array methods
  -Nonlinear Regression methods
  -Manifold Embedding methods
  -Flexible Models
  -Geometric methods
  -Tracking methods
  -Hybrid methods

Appearance Template methods

- Uses image based comparison metrics.
- Compares new image with a set of exemplars.
- Advantages:
  -dataset can be easily expanded.
  -negative training examples and facial feature points are not required.
  -well suited for low and high resolution imagery

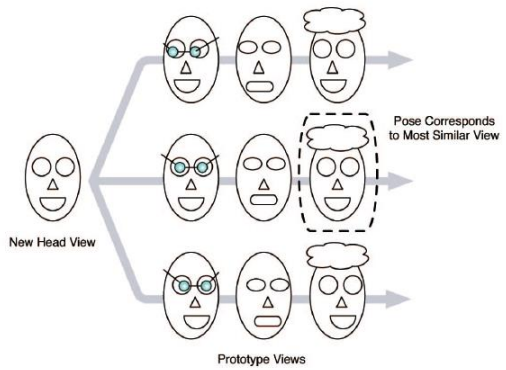

Fig. 3. **Appearance template methods** compare a new head view to a set of training examples (each labeled with a discrete pose) and find the most similar view.

Appearance Template methods

- Disadvantages:
  -estimates only discrete poses.
  -depends on reliable head region detection.
  -computationally expensive for large datasets.
  -pair wise similarity does not necessarily mean pose similarity.

Detector Array methods

- Multiple trained face detectors.
- Each trained to a different discrete pose.
- Similar to Appearance Template methods.
- Advantages:
  -Separate head detection not required.
  -training algorithms ignore variations in appearance.
  - well suited for low and high resolution imagery

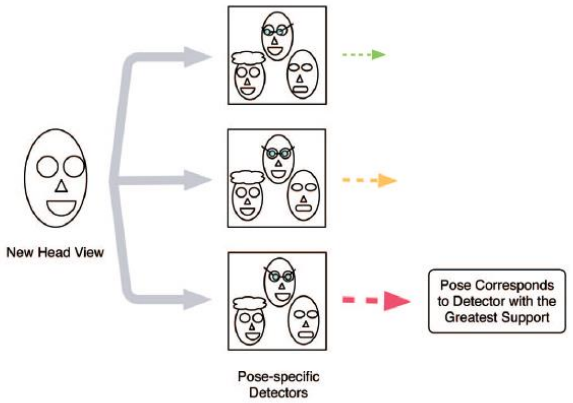

Fig. 4. **Detector arrays** are comprised of a series of head detectors, each attuned to a specific pose, and a discrete pose is assigned to the detector with the greatest support.

Detector Array methods

- Disadvantages:
  -computationally expensive.
  -burdensome to train multiple detectors.
  -more training data is required.
  -positive and negative samples could be quite similar.
  -ambiguity if more than one detector classifies the image as positive.

Nonlinear Regression methods

- Nonlinear mapping from Image space to pose space.
- A model is built based on a labeled training data.
- Some of the techniques used: SVRs, MLP, LLM PCA is used to reduce dimensionality.
- Alternatively, facial feature locations can also be used.

Nonlinear Regression methods

- Advantages:
  -fast and most accurate.
  -requires just the cropped labeled faces for training.
- Disadvantage:
  -prone to error from poor head localization.

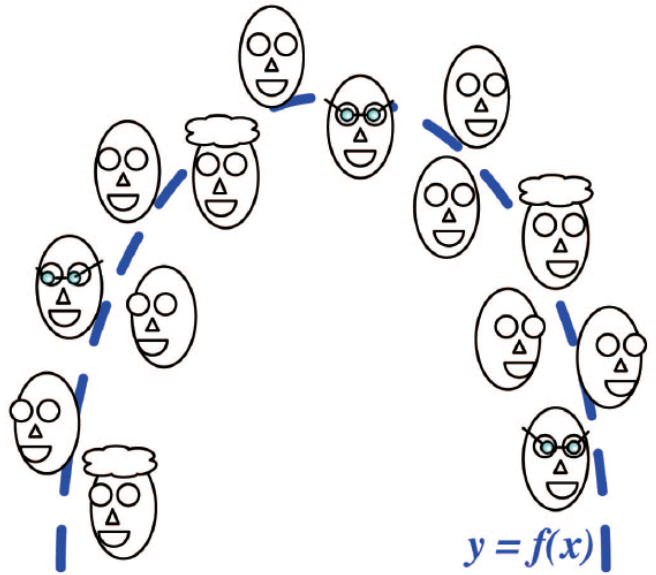

Fig. 5. **Nonlinear regression** provides a functional mapping from the image or feature data to a head pose measurement.

Manifold Embedding methods

- Treats each image as a sample in high dimensional space and to be lying on a low dimensional continuous manifold.
- The manifold is modeled and an embedding technique is used to map a new sample into the manifold.
- Regression or template matching can then be used to estimate pose.
- Manifold mapping techniques: PCA, KPCA, LDA, LLE etc

Manifold Embedding methods

- Advantage:
  -embedding can be performed by simple matrix multiplication
- Disadvantage:
  -lack representational ability of the non-linear techniques.

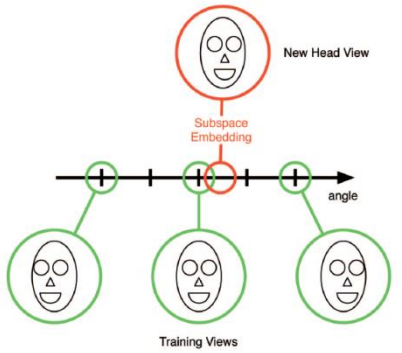

Fig. 6. **Manifold embedding methods** try to directly project a processed image onto the head pose manifold using linear and nonlinear subspace techniques.

Flexible models

- Use a non-rigid model which is iteratively fit to conform to the facial structure of each individual.
- Uses training data with annotated facial features.
- Examples: EGM, AAM.

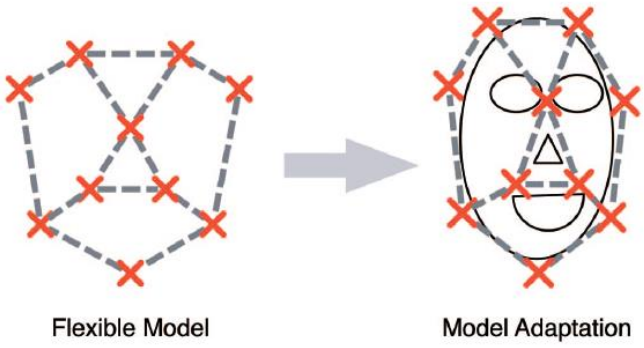

Fig. 7. **Flexible models** are fit to the facial structure of the individual in the image plane. Head pose is estimated from feature-level comparisons or from the instantiation of the model parameters.

Flexible models

- Advantages:
  -Invariance to head localization error.
  -less inter-subject variability.
- Disadvantages:
  -computationally expensive
  -does not work well with low resolution images.

Geometric methods

- Use precise configuration of local features.
- Advantages:
  -fast and simple.
- Disadvantages:
  -require high precision location of the features.
  -does not work with low resolution images.
  -occlusion

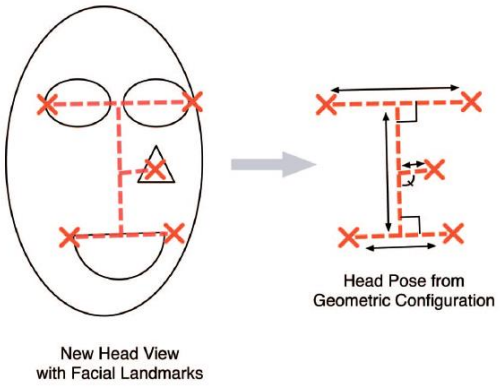

Fig. 8. **Geometric methods** attempt to find local features, such as the eyes, mouth, and nose tip, and determine pose from the relative configuration of these features.

Tracking methods

- Uses temporal continuity to provide the head pose estimate by tracking the head.
- Initial position of the head needs to initialized.

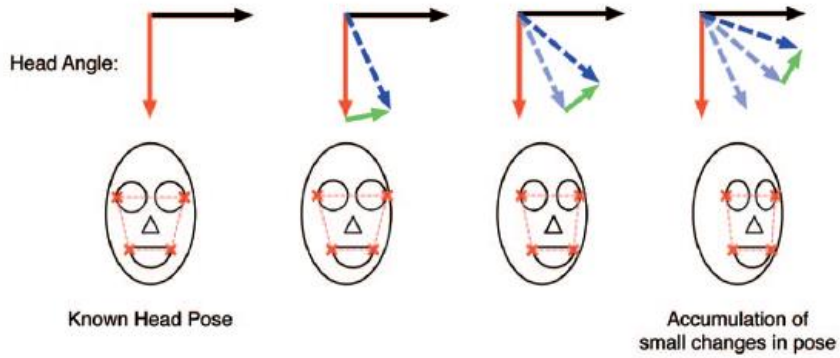

Fig. 9. **Tracking methods** find the relative movement between video frames to estimate the global movement of a head.

Tracking methods

- Advantages:
  -Very accurate tracking is available.
  -Some of the challenges are taken care by the tracking algorithm itself.
- Disadvantages:
  -Initialization needs to be very accurate.
  -Semi automatic.

Hybrid methods

- Combination of the other methods.
- Advantage - Overcome limitations of the individual methods.
- Disadvantage - Resulting algorithm could be very complex.

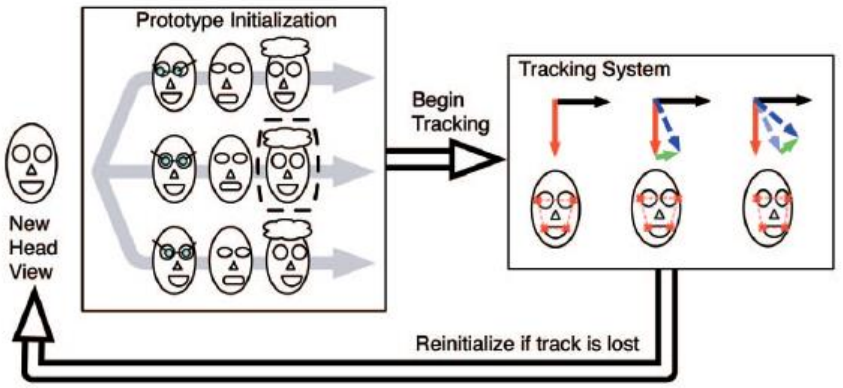

Fig. 10. **Hybrid methods** combine one or more approaches to estimate pose. This image is an example of an appearance template method combined with a point tracking system.

Ground Truth Datasets

- Necessary to evaluate and compare the performance of head pose estimation systems.
- Methods to obtain ground truth data:
  -Directional suggestion
  -Directional suggestion with laser pointer
  -Manual annotation
  -Camera arrays
  -Magnetic sensors and Inertial sensors
  -Optical motion capture systems

Questions?
Thank You!

Questions

- Write a note on Detector Array methods and Tracking methods.
- How are the detector array methods different from Appearance Template methods?
- Write a brief note on any two methods for capturing ground truth data for head pose estimation.
- Which method for ground truth data acquisition is the best in your opinion? State your reasons.