3D Face Reconstruction from Stereo Video

Unsang Park
Advisor: Anil K. Jain

Biometrics Research Group
Michigan State University
biometrics.cse.msu.edu

Contents

- Motivation
- 3D face reconstruction from video
- Face recognition with reconstructed 3D face models
- Automatic facial landmark detection and reconstruction
- Conclusions and future Work

Motivation

- The same face in a video undergoes substantial variations in pose, illumination, etc. → frontal face recognition does not work
- Videos captured by surveillance systems cannot be used for subject identification

Washington Dulles airport security video of the "hijackers" of 9/11 (AP News).
Surveillance system does not provide the capability of subject identification

3D Face Reconstruction from Video

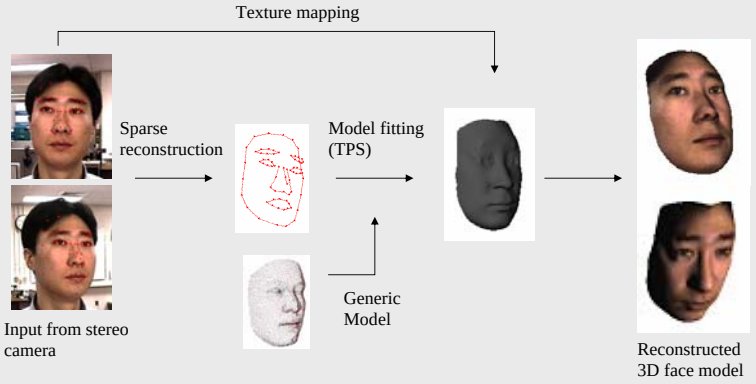
- 3D Face reconstruction from Video
  - Pose, lighting and expression can be compensated for better 2D face recognition
  - Verification/Identification in 3D domain using 2D sensors

SfM vs. Stereography

- SfM (Structure from Motion)
  - Finding R, T, X from n-view geometry given a set of corresponding points
- Stereography
  - Finding X from 2-view geometry given R, T and a set of corresponding points
  - More accurate than SfM
  - Limited applications
    - R: rotation matrix
    - T: translation matrix
    - X: 3D world coordinates

3D Face Reconstruction from Stereo Video

- Overall approach

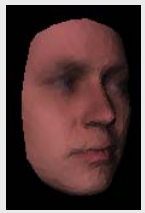


Generic Model Construction

- Semi-dense generic model is constructed from Blanz's morphable model

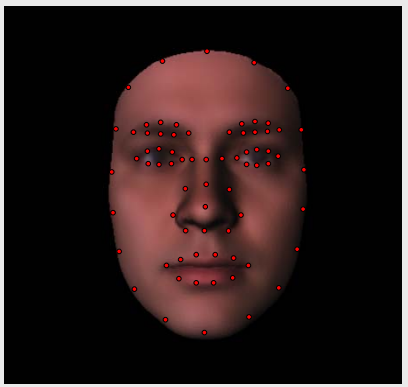
Original morphable model (~70000 vertexes)

5000 vertices

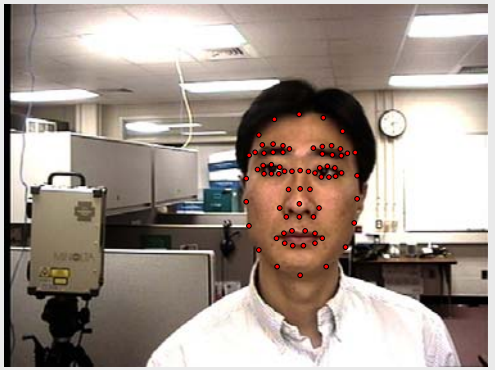


500 vertices

Feature points labeling



72 feature points on the generic model



72 feature points on the video frame

Sparse Reconstruction with Stereography

- Back projection to 3D space

$$x = PX \qquad P = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix}$$

$$w\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} X \qquad \begin{bmatrix} P_3 x - P_1 \\ P_3 y - P_2 \end{bmatrix} X = 0$$

$$\begin{bmatrix} P_3 x - P_1 \\ P_3 y - P_2 \\ P_3' x' - P_1' \\ P_3' y' - P_2' \end{bmatrix} X = 0$$

Dense Reconstruction with Thin Plate Spline

- Calculates the deformation mapping function from a set of control points *u* to *v*.

F(u) = c + Au + W^Ts(u), c: translation, A: rotation

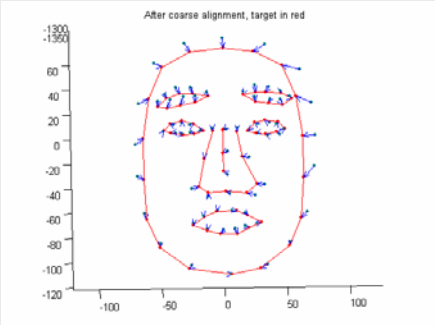
W: non-linear deformation

$$s(u) = (\sigma(u\text{-}u_1), \sigma(u\text{-}u_2), \cdots, \sigma(u\text{-}u_n))^T, \qquad \sigma(u) = \begin{cases} \|u\|^2 \log(\|u\|), & \|u\| > 0 \\ 0, & \|u\| = 0, \end{cases}$$

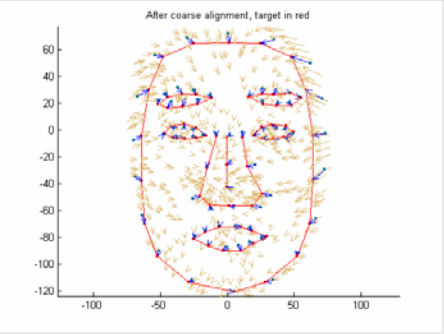
$$\begin{bmatrix} s & 1_n & u^T \\ 1_n^T & 0 & 0 \\ u & 0 & 0 \end{bmatrix} \begin{bmatrix} W \\ c^T \\ A^T \end{bmatrix} = \begin{bmatrix} v \\ 0 \\ 0 \end{bmatrix}$$

Dense Reconstruction with Thin Plate Spline

- Example TPS fitting process



Fitting of 72 feature points



Fitting of the rest points (5000) according to the TPS deformation function

Reconstruction Results

Input images

Reconstructed 3D models

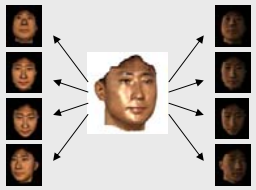
Real 3D models

3D model Assisted Face Recognition In Video

- All possible pose and illumination variant images of a subject are not usually available
- Pose and illumination variations can be generated from the 3D face models
- 2D image based face matcher are readily available (e.g., FaceIt® or FaceVAC®)

Probe images (from a video)

2D to 2D matching

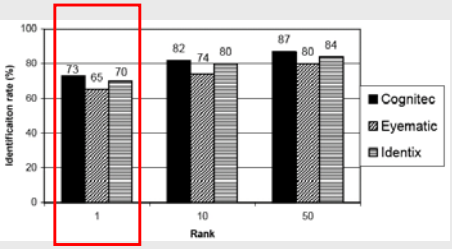


Gallery images (from a 3D model)

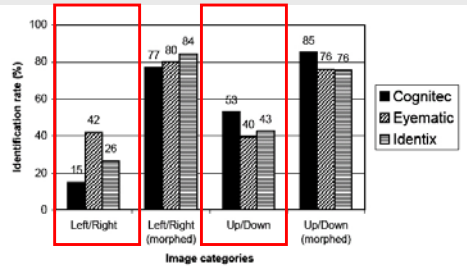
3D model Assisted Face Recognition In Video

- Baseline performance of FaceIt (Identix) and FaceVAC (Cognitec) from FRVT 2002

Example probe images



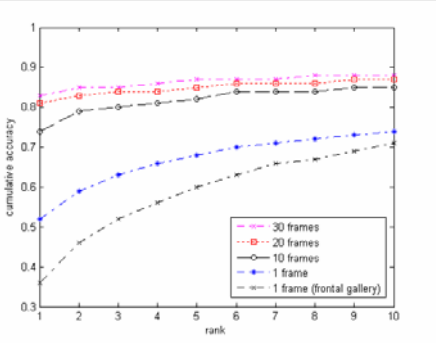
Rank-1 identification accuracy, 37437 subjects



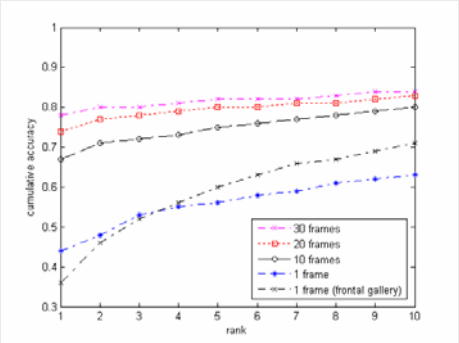
Performance degradation with pose variations, 87 subjects

Face Recognition with Reconstruction

- 15 subjects in probe and 100 subjects in gallery
- Score-sum fusion of the matching score from FaceIt® and FaceVAC® by Min-Max normalization
- Frame level fusion is performed by majority-vote

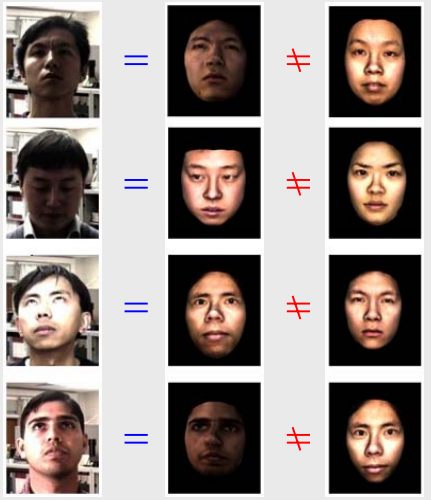


Face recognition using 3D models acquired from range sensor



Face recognition using reconstructed 3D models

Matching Examples



Probe

Correct matching with pose/lighting variation

Incorrect matching without pose/lighting variation

Automatic Facial Landmark Detection

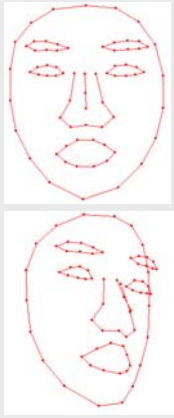
- 72 landmarks using Active Shape Model (ASM) on a Video with 60 frames

Landmark detection without temporal coherency

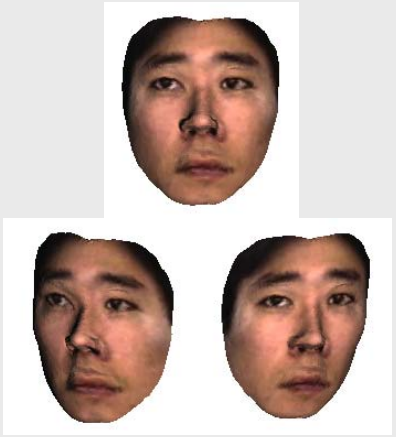
Landmark detection with temporal coherency (estimated feature points at current frame are used as the initial state for the next frame)

Reconstruction with SfM

- Reconstruction is performed using automatically detected/tracked 72 landmarks on 60 frames by factorization method



Sparse reconstruction



Dense reconstruction with texture mapping

Conclusions

- A (dense) 3D face reconstruction method using stereography, TPS and generic model is proposed
- Reconstructed 3D face model provides better face recognition performance by adding pose/lighting variations into gallery
- Preliminary work on automatic facial landmark detection using ASM, followed by SfM-based reconstruction shows promising results

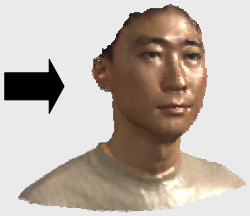
Future work

- Developing more robust 3D face reconstruction with SfM
- Real-time 3D face reconstruction from video
- Building a frame work of using reconstructed face models from a video in face recognition task (selecting multiple frames incrementally for the 3D reconstruction and recognition)
- Automatic facial pose/lighting estimation in video for the 3D face reconstruction and recognition

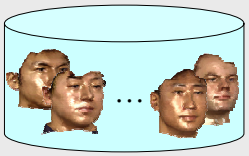
3D Face Model Construction

- 3D face models are used to generate synthetic 2D face images at various pose and lighting

Multi-view 3D face scanning
With Minolta VIVID 910 scanner



Full 3D model



3D database