Visual Sensing of 3D Human Body for Immersive Interactions

Isaac Cohen
icohen@usc.edu

Institute for Robotics and Intelligent Systems
Integrated Media Systems Center

University of Southern California

Perceptual User Interface

- Natural human computer interaction (HCI)
  - Visual sensing of user
  - Unencumbered human, natural background
- Interaction with synthetic environment
  - 3D pose estimation for adaptive rendering
  - Gesture recognition for augmented, natural communication
- Applications
  - Computer-Aided Training: tracking trainees/expert
  - Virtual prototyping

Perceptual User Interface: Example

Minority Report Trailer

Perceptual User Interface: Overview

- Interaction with a synthetic environment
  - Gesture recognition for augmented, natural communication
  - Identifying user's body and facial gestures
- Visual sensing of user in interactive environment
  - Head pose tracking
  - Full body / hands and fingers tracking
- Applications
  - Computer-aided training: tracking trainees/expert
  - Interaction with virtual objects
  - Virtual prototyping

Perceptual User Interface: Head Pose Tracking

- Automatic face detection
  - PCA/ICA face features
  - SVM-based detection
- Face pose estimation
  - Head modeling using 3D cylinder
  - Fitting 3D model to 2D image features
  - Recovering 3D rigid motion parameters

Face Detection: Machine Learning Approach

- Model faces features
  - Color, Skin, Textures
    - Principal component analysis: Eigenfaces
    - Independent component analysis
  - Face geometry
    - Detecting face parts and their relationships
- Classification of features
  - Support Vector Machine

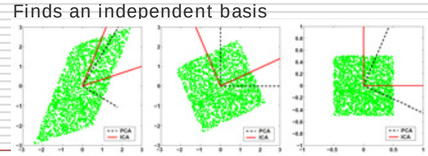
Extracting Face Features: Independent Component Analysis

- Problem:
  - Assume that we observe linear mixtures $x_1, x_2 \dots x_n$ of n independent components $s_1, s_2 \dots s_n$:

    $x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n$ $\qquad x = As$
  - Estimate the matrix A and $s_1, s_2 \dots s_n$
- Solution
  - Independent components are found by looking for directions in which the data has a non-Gaussian distribution
  - The estimate of *s* is *u* defined by $u = Wx$ and W is the inverse of A

ICA vs. PCA

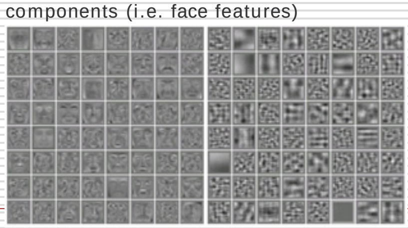
- PCA (Principal Component Analysis)
  - Find a basis that maximizes the variance of the projected data: uncorrelated basis
  - Main application: dimension reduction of the data
- ICA
  - Finds an independent basis



(a) Original (b) After PCA pre-whitening (c) After ICA projection

Face Features Extraction

- Given a large set of face images
  - Use PCA for dimensionality reduction
  - Use ICA for characterizing independent components (i.e. face features)



Faces Non-Faces

Features Classification: Support Vector Machine

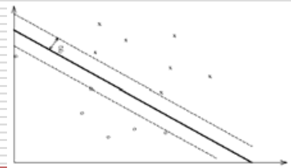
- Definition of a decision function from a set of observations
- Given a set of observations x with corresponding label (i.e. face/non-face)
- Characterize the best hyper-plane separating the observation into 2 classes
- Decision function:

$$f(x) = \operatorname{sgn}\left(\sum_{i=0}^{l-1} \alpha_i^0 y_i K(x_i, x) + g\right)$$

α_i^0 maximizes

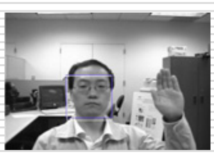
$$W(\alpha) = \sum_{i=0}^{l-1} \alpha_i - \frac{1}{2}\sum_{i,j=0}^{l-1} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

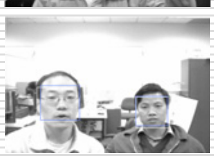
K kernels



Face Detection

- Scan the image with a fixed/variable size neighborhood
- Estimate the corresponding independent component (multiply by W)
- Evaluate the SVM decision function f
- If f>0, the region correspond to a face



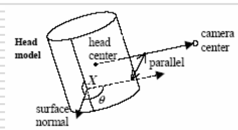


Face Detection: Summary

- Limitation:
  - Dependent on the set of training data:
    - Face orientation
    - Illumination
  - Selection of neighborhood
- Advantage
  - Fronto-parallel views
  - Fast

Head Pose Estimation

- Problem: Characterize the pose of the person. i.e. determine the rotation and translation in 3D
  - Head modeling using 3D cylinder
  - Fitting 3D model to 2D image features
  - Recovering 3D rigid motion parameters from computed 2D motion



Head Pose Estimation

$I(u,t)$ Image at time t and pixel u=(x,y)

At time t+1 and the pixel u=(x,y) moves to $u' = F(u,\mu)$

where $\mu = (w_x, w_y, w_z, t_x, t_y, t_z)$ are the motion parameters

and $F(u,\mu)$ is the motion model

Assuming that the illumination condition do not change

$$I(F(u,\mu), t+1) = I(u,t)$$

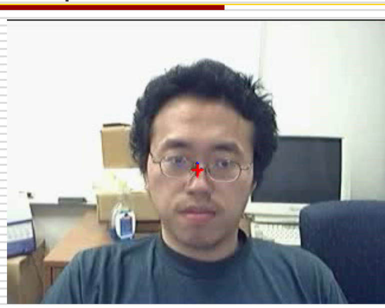
The motion parameters can be characterized by minimizing

$$E(\mu) = \sum_{u\in\Omega} (I(F(u,\mu), t+1) - I(u,t))^2$$

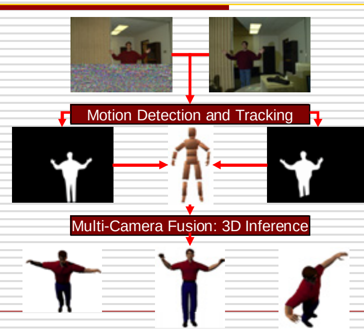
The solution is obtained by Euler-Lagrange

$$\left(\sum_{u\in\Omega} (I_u F_\mu)^T (I_u F_\mu)\right)\mu = \sum_{u\in\Omega} I_t (I_u F_\mu)^T$$

Example 2

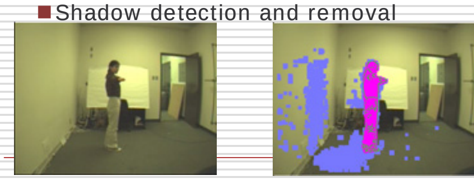


3D Body Modeling: Overview



Silhouettes Detection

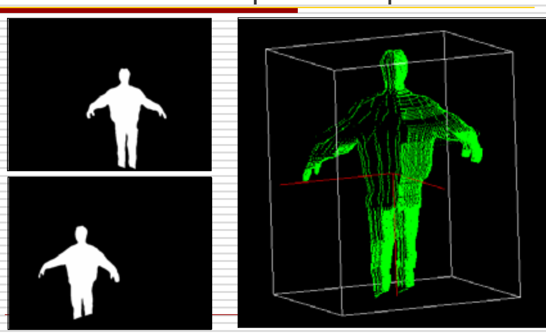
- Foreground detection:
  - Gaussian background modeling
  - Edge detection
  - Distance transform of binary image
  - Shadow detection and removal



Foreground detection

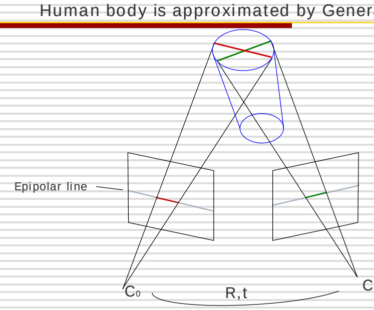
- Construct an adaptive model of "background"
  - Each pixel modeled as a multi-dimensional Gaussian distribution in color space
  - Updated when new pixel values are observed
- Extract "foreground"
  - A pixel is foreground if sufficiently different than current background model
  - Marked pixels grouped into connected components

3D Body Modeling: GC-based Shape Description



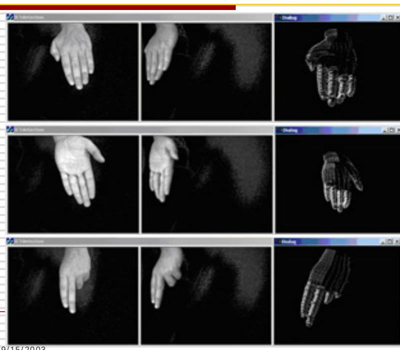
Why does it work?

Human body is approximated by Generalized Cylinders



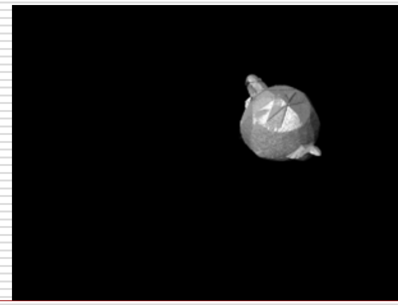
Epipolar Geometry

3D Hand Modeling: GC-based Reconstruction



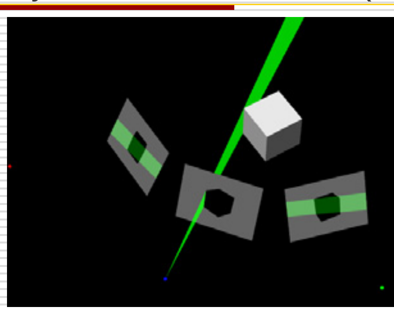
I. Cohen – CS597 – 09/15/2003

Using more than 2 cameras: Visual Hulls



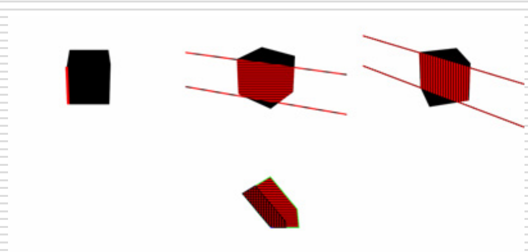
I. Cohen – CS597 – 09/15/2003

Polyhedral Visual Hulls (PVH)



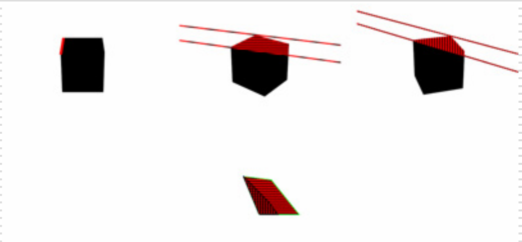
I. Cohen – CS597 – 09/15/2003

Computing the PVH

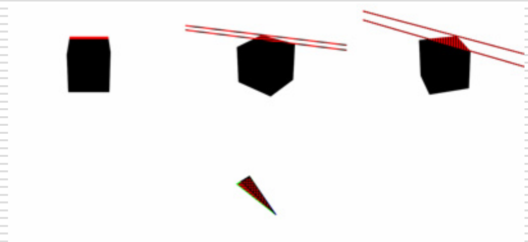


I. Cohen – CS597 – 09/15/2003

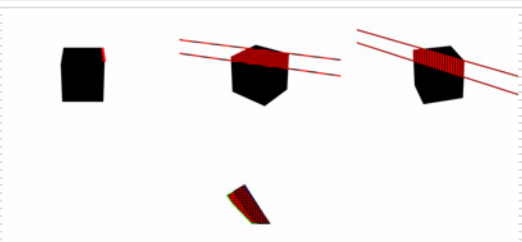
Computing the PVH



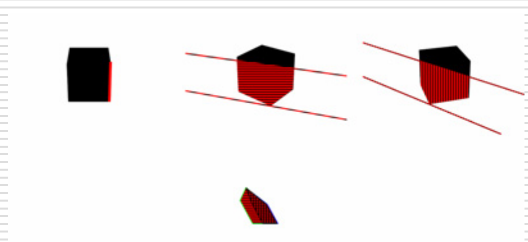
Computing the PVH



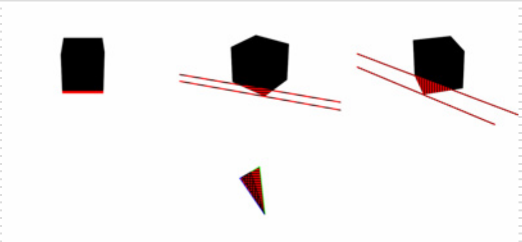
Computing the PVH



Computing the PVH

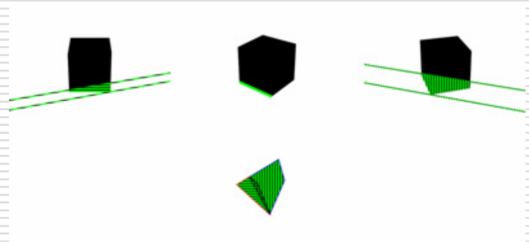


Computing the PVH



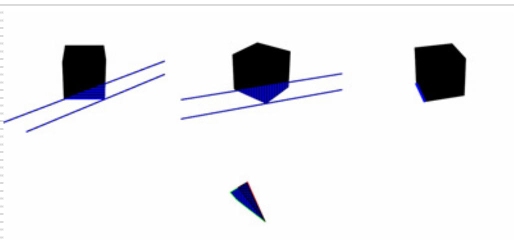
I. Cohen – CS597 – 09/15/2003

Computing the PVH



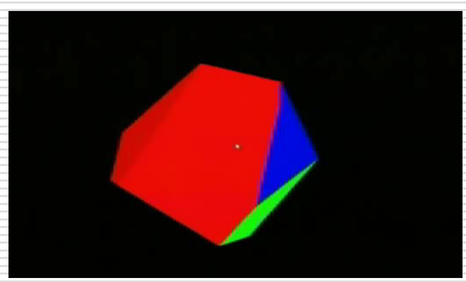
I. Cohen – CS597 – 09/15/2003

Computing the PVH



I. Cohen – CS597 – 09/15/2003

Final Result



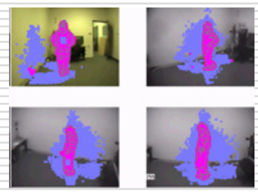
I. Cohen – CS597 – 09/15/2003

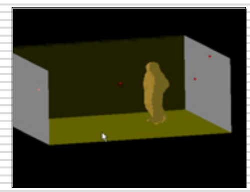
Human Posture Recognition

- Understanding human actions
- Inference of human body postures
  - Use of 3D Visual-Hull reconstruction
  - Support Vector Machine
    - Learning of human postures
    - Identification of currently 13 body postures
- Describing generic gestures
  - Use of postures combinations

Visual Hull Reconstruction

- Polygonal approximation of silhouettes
- Polyhedral visual hull reconstruction
- Human body Visual Hull reconstruction



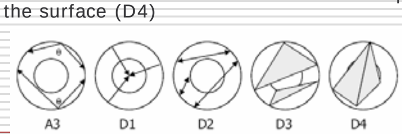


Global Shape Descriptors

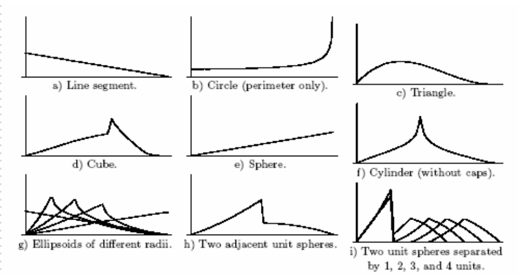
- Characterizing the shape of a 3D surface
  - Querying 3D surface databases
  - Measuring similarity between surfaces
- Wish list:
  - Invariance to deformations
    - Translations, Rotation, Scale
    - Affine invariance?
    - Projective invariance?
  - Localized variations
    - Local change in the data generate a local change in the surface descriptor

Selecting a Shape Function: Geometric Measurements

- Angles between 3 random points on a surface of a 3D model (A3)
- Distance between a fixed point and one random point on the surface (D1)
- Distance between 2 random points on the surface (D3)
- Area of the triangle between 3 random points (D3)
- Volume of a tetrahedron between 4 random points of the surface (D4)

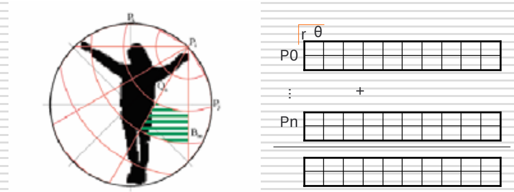


D2 - Descriptor



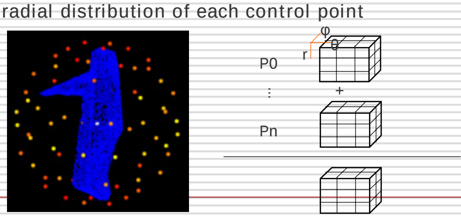
Global Shape Description in 2D

- 2D shape descriptor
  - For every control point we accumulate in each bin the number of silhouette points



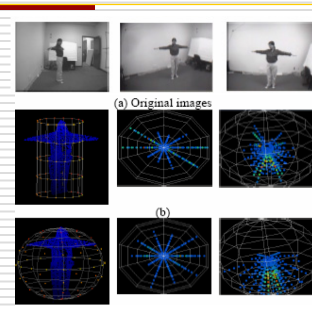
Global Shape Description in 3D

- Generalization to 3D
  - Define a sphere containing the visual hull
  - Sample the sphere using a set of control points
  - Derive the global shape description from the radial distribution of each control point

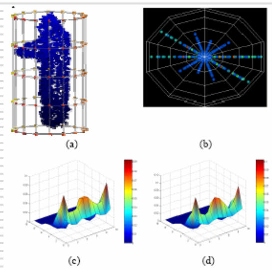


Selection of a Reference Surface

- Sphere:
  - 3D symmetry
- Cylinder:
  - axial symmetry



Localized Variations



Measuring Shapes Similarity

- Chi2: $D(f,g)=\int\frac{(f-g)^2}{f+g}$
- Bhattacharyya: $D(f,g)=1-\int\sqrt{fg}$
- Minkowski $D(f,g)=\left(\int|f-g|^N\right)^{1/N}$
- Kullback-Leibler $D(f,g)=\int f\log\frac{f}{g}$

Shape Description: Rotation Invariance

		Posture 1					Posture 3				
		0	16	32	48	60	0	16	32	48	60
Posture 1	0	0.000000	0.013796	0.016846	0.015525	0.013204	0.061394	0.054847	0.062393	0.058276	0.056100
	16	0.014024	0.000000	0.015648	0.014420	0.019950	0.066901	0.057760	0.065960	0.063246	0.059444
	32	0.015716	0.014020	0.000000	0.012110	0.015999	0.058286	0.048821	0.058448	0.058678	0.054301
	48	0.015510	0.014088	0.012906	0.000000	0.012449	0.060668	0.050619	0.058655	0.057175	0.056948
	60	0.012907	0.018557	0.016216	0.012712	0.000000	0.062684	0.049774	0.058670	0.057069	0.055565
Posture 3	0	0.061622	0.061855	0.059343	0.062843	0.064294	0.000000	0.014499	0.017765	0.015684	0.015935
	16	0.053709	0.053193	0.048824	0.052622	0.051121	0.014329	0.000000	0.011320	0.014403	0.013560
	32	0.059374	0.059499	0.055958	0.059888	0.058492	0.016808	0.010753	0.000000	0.010592	0.012479
	48	0.056869	0.057757	0.056807	0.057576	0.056641	0.015286	0.013643	0.010623	0.000000	0.009813
	60	0.056294	0.055608	0.054268	0.058379	0.056175	0.015390	0.012993	0.012427	0.009417	0.000000

•**Kullback-Leibler Distance or entropy**

Shape Description: Scale Invariance

Posture 1	1	0.9	0.9^2	0.9^3	0.9^4
1	0.000000	0.000000	0.000000	0.000000	0.000000
0.9	0.000000	0.000000	0.000000	0.000000	0.000000
0.9^2	0.000000	0.000000	0.000000	0.000000	0.000000
0.9^3	0.000000	0.000000	0.000000	0.000000	0.000000
0.9^4	0.000000	0.000000	0.000000	0.000000	0.000000

•**Kullback-Leibler Distance** (*scale rate is 0.90*).

Support Vector Machines

- SVM-based Classification
  - Duality
  - Operate in a Kernel induced feature space
  - Maximize Margin
  - Convexity: no local minima
  - Sparseness
- 3D shape are encoded as vectors:

posture ID bin-index:weight bin-index:weight

Experimental results: Defined Postures

Posture Recognition

Future Goals

- Feedback to 3D-CAD data
- Interactive visualization
- Distributed Interaction
- Human behavioral modeling:
  - multiple users
- Computer Aided Training
- Occupational therapy