3D Projective Shape Analysis in Face Recognition from Digital Camera Images

Ka Chun Wong
Florida State University

April 25, 2022

Outline I

Processing

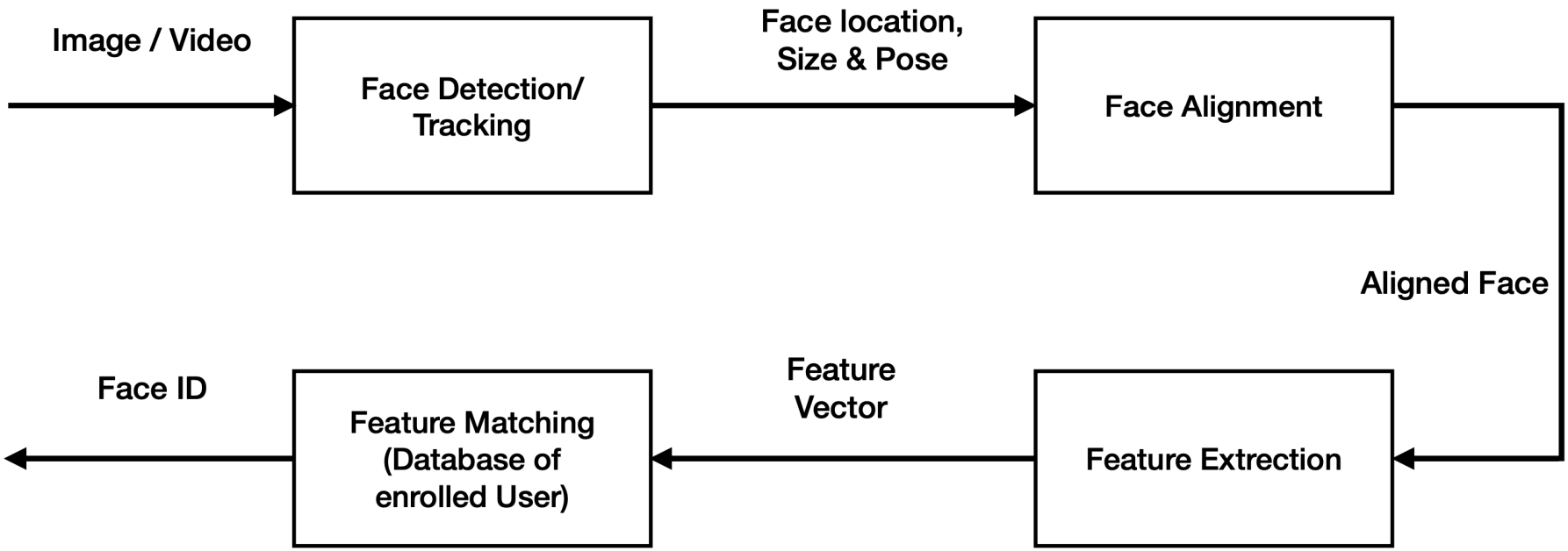


Figure: Face Recognition Precess

Challenges

- Large Variability in Facial Appearance
  - Shape, Pose, Facial expression
  - Illumination, Camera
  - Face-Based Person Identification
  - Variation by illumination > change in face identity
- High Dimensionality and Small Sample Size
- Faces are Highly Complex nonconvex surfaces

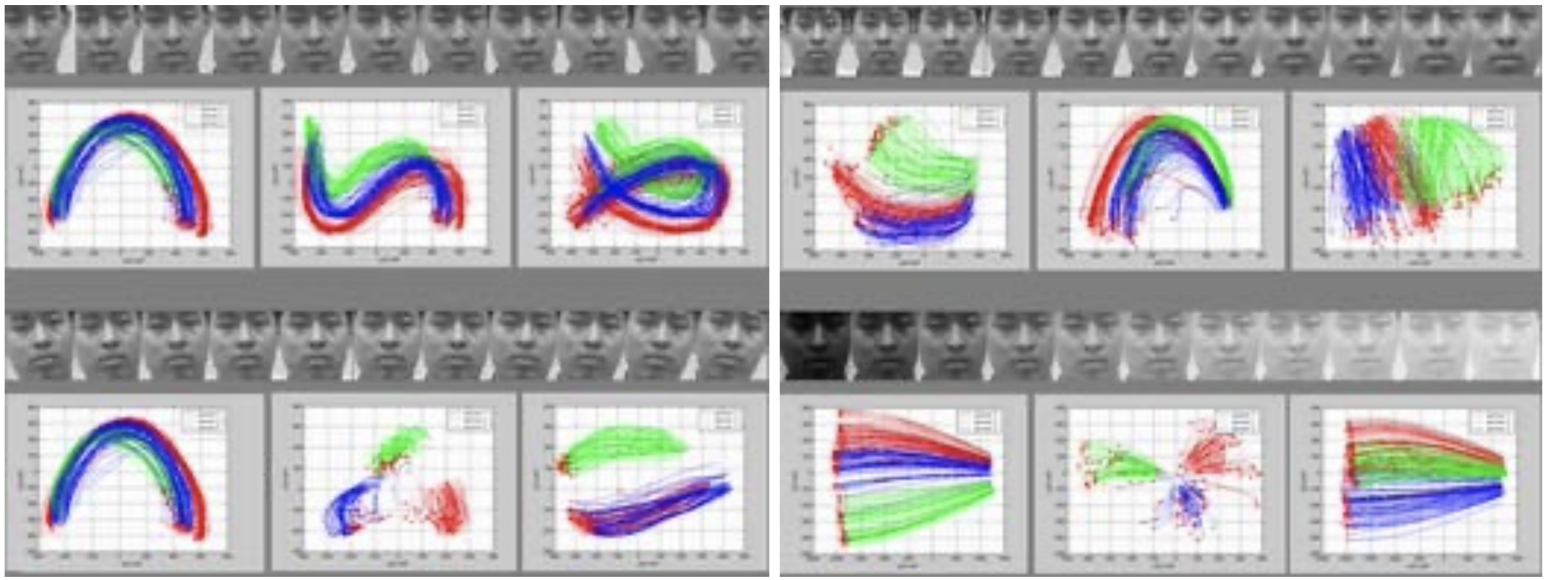


Figure: Nonlinearity and nonconvexity of face manifolds under translation, rotation , scaling, and pointwise lighting change.

Short Brief on Detection

- Appearance-Based approaches & Learning-Based approaches
  - PCA, Support Vector Machine, Adaboost ...
- Adaboost : a complex nonlinear strong classifier $H_N(x)$ is constructed as a linear combination of N simpler, easily constructible weak classifiers in the following form

$$H_N(x) = \frac{\sum_{n=1}^{N} \alpha_n h_n(x)}{\sum_{n=1}^{N} \alpha_n} \tag{1}$$

where x is a pattern to be classified, $h_n(x) \in \{-1, +1\}$ are the N weak classifiers, $\alpha_n \geq 0$ are the combining coefficients in $\mathbb{R}$

Figure: Appearance-Based approach

Short Brief on Detection

- Preprocessing : Skin Color Filtering, Image Normalization

Figure: Skin Color Filtering

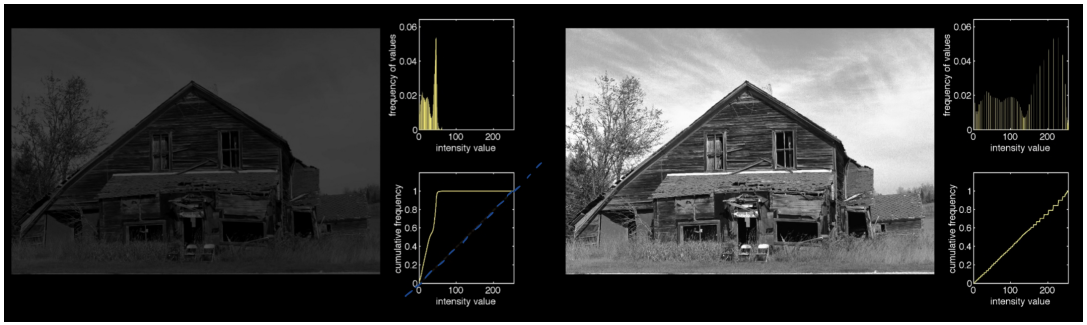
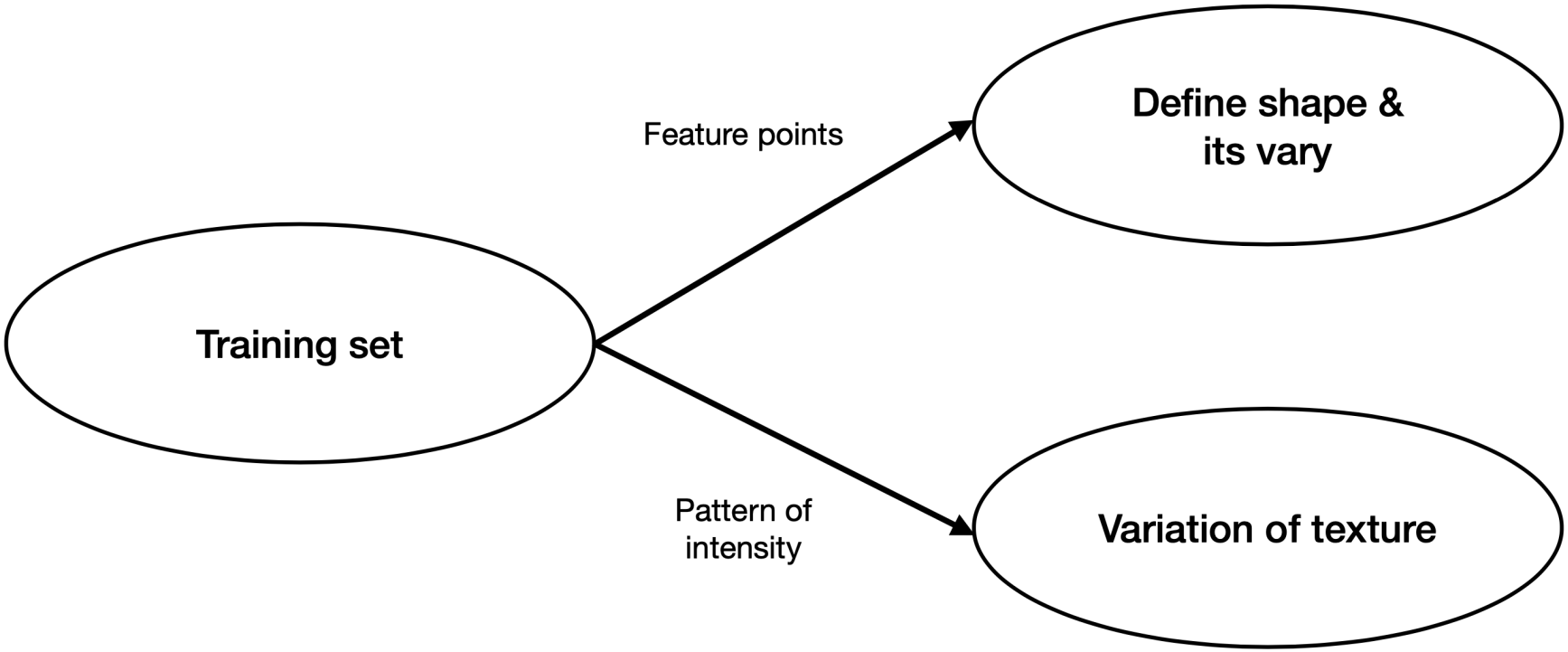


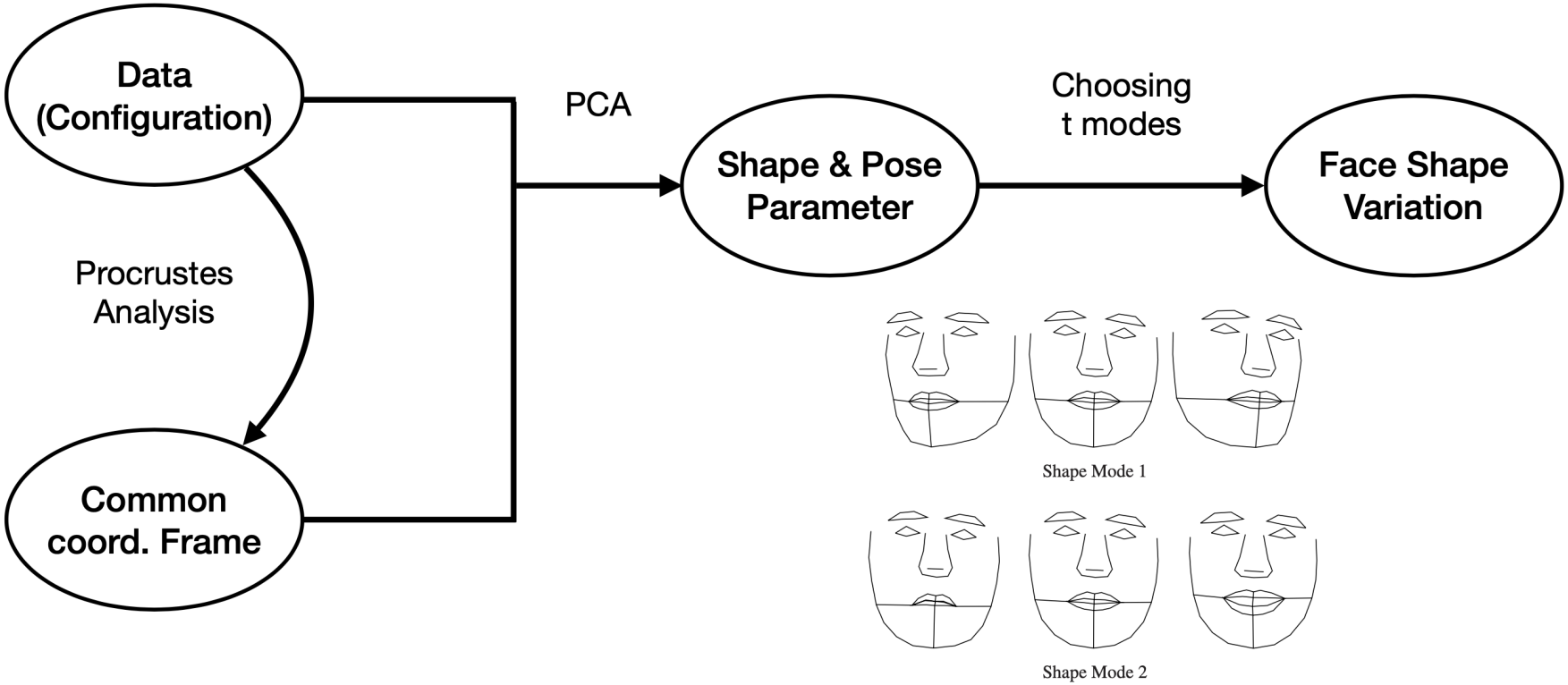
Figure: Histogram Equalization : a way for Image Normalization

- Head rotation : in plane, out of plane, up and down
- Postprocessing : Merging multiple detection, Elimination on false alarms

Facial Modeling

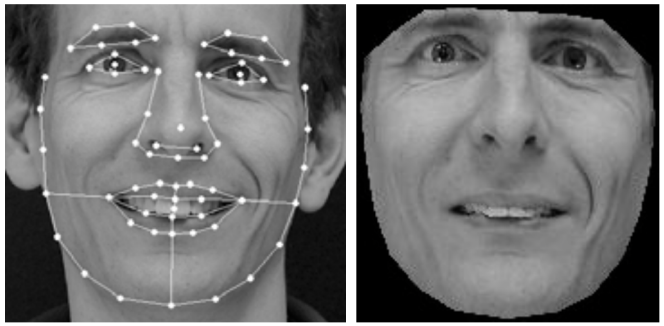


Statistical Model of Shape



Statistical Model of Texture & Appearance

- Warp the image by a warping function so the feature points match the reference frame



Original image Warped to mean shape

- Normalized to remove global lighting effect by linear transformation
- Evaluated the texture parameter and modes by PCA
- Combining shape and texture model, Analysing the correlation between shape and texture, Resulting in a Appearance Model

Active Shape Model

- Modify the parameters to better fit the model to a new image.
- Find the new position along the normal of the model point.

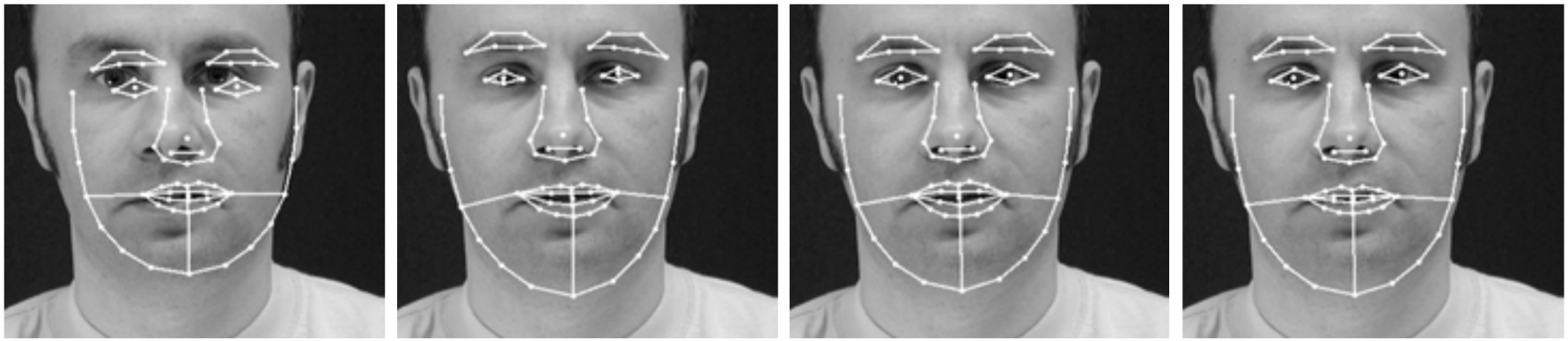


Figure: Active Shape Model

Active Appearance Model

- Given a new image, locate the facial features and extract information
- Generate a face by the model and minimize the difference to the image
- Locate the position of features by the model parameters from the generated face

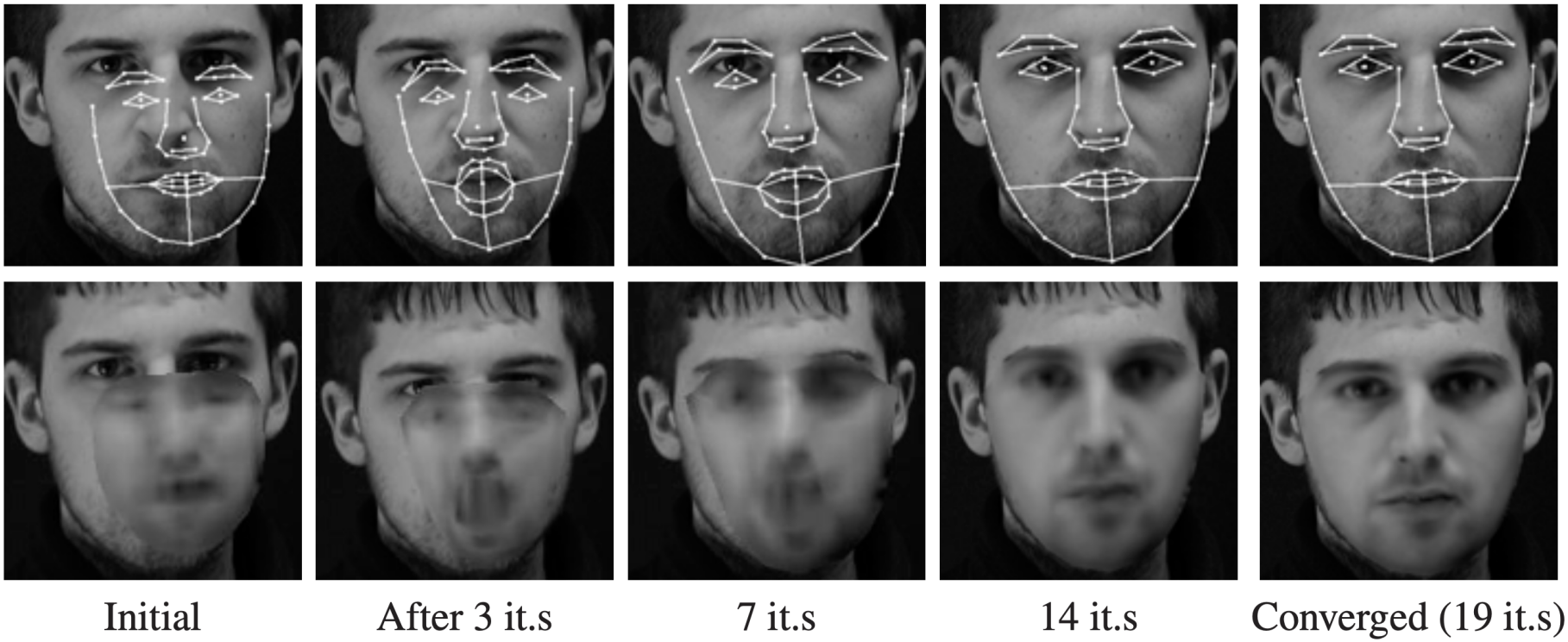
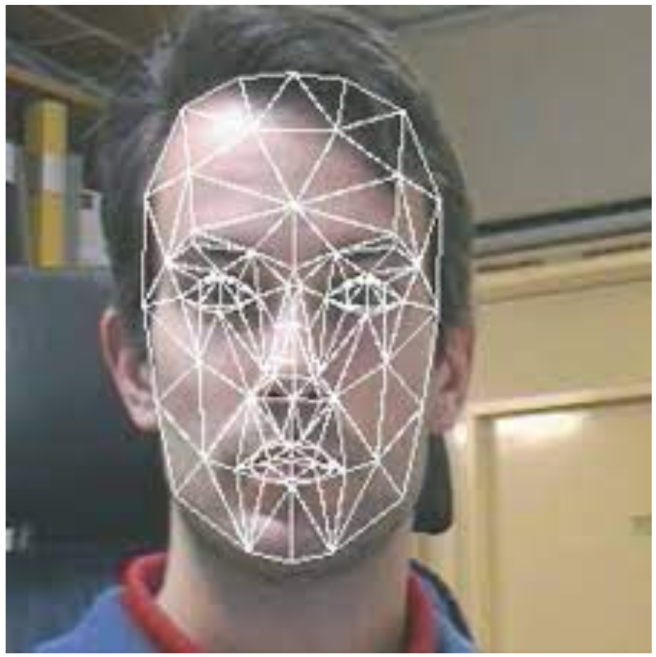


Figure: Active Appearance Model

Example of Parametric Face Modeling

- Eigenface
- Facial Action Coding System
- MPEG-4 Facial Animation
- Computer Graphics Models : Wireframe model
- Candide



Illumination and Skin Color

- Light will affect appearance of face
- Cameras are not able to distinguish changes of surface colors from color shift caused by varying illumination spectra
- Color space : RGB, HSV, HSL ...
- Intensity vs. Chromaticity
- Chromaticity providing robustness against changing intensity
- Skin tones differ mainly in intensity

Illumination and Skin Color

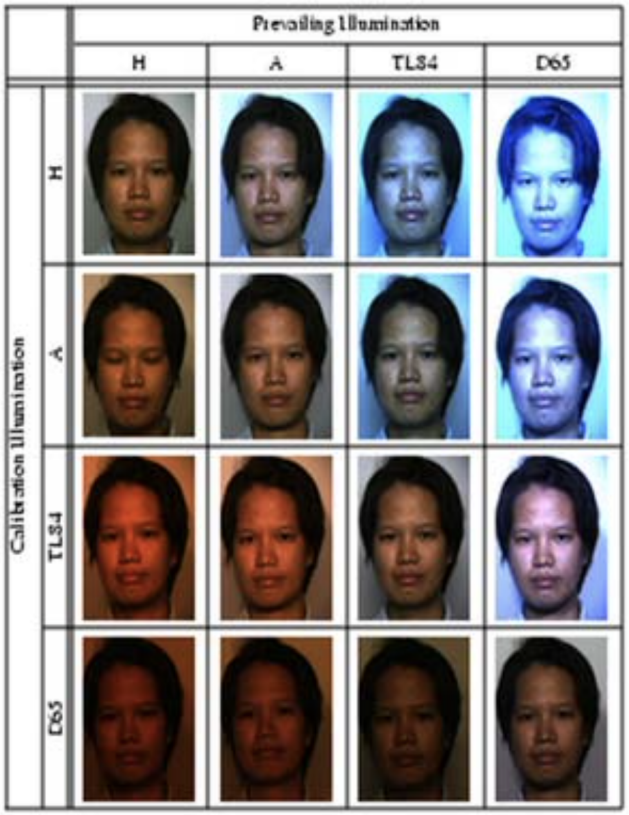
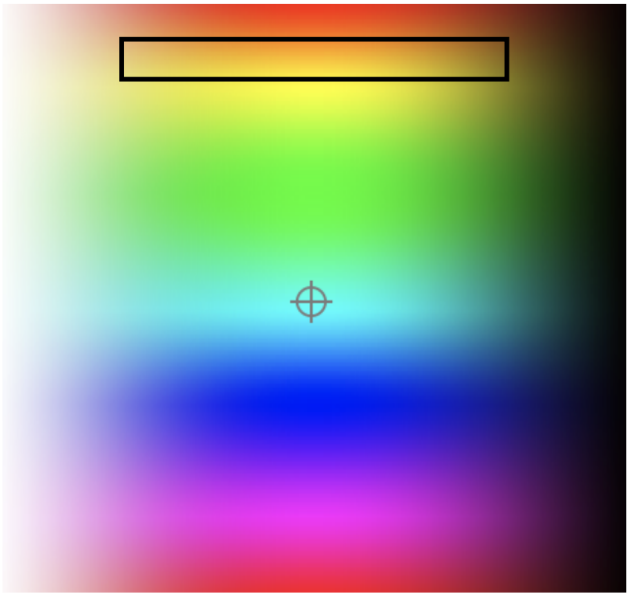


Figure: A 16 image series has been obtained with four white balancing conditions (rows) and four prevailing illumination options (columns).

From the Statistics perspective though, such data as the images above are irrelevant, as they are obtained for a singleton (sample size $n = 1$).

Skin Color Model

- Probability assigned to the colors of different object classes
  - Certain color tones belong to an object's pixels more often than others
- Spatial limitation of object's colors in a color space
  - Find a color space region in which the object color pixels typically fall

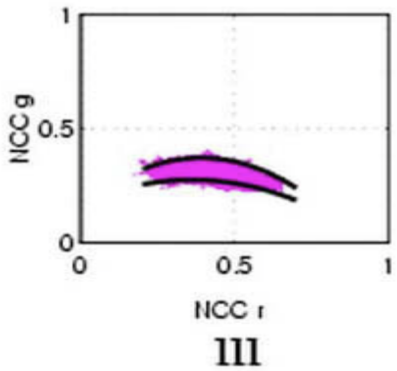


Skin Locus

1. Collecting the image set for different camera calibrations and prevailing conditions.
2. Selecting the skin areas for further processing.
3. Converting the selected skin RGBs into a chromaticity space and modeling them using a polynomial.

I

II



III

Projective Geometry

- Given a real vector space V, the projective space $P(V) = \{[x], x \in V \setminus 0_V\}$
- $\mathbb{R}P^m \doteq P\mathbb{R}^{m+1}$
- A projective point $p = [x] = [x^1 : x^2 : \cdots : x^{m+1}]$ featuring the homogeneous coordinates $(x^1, x^2, \ldots, x^{m+1})$ of p, which are determined up to a multiplicative constant
- A projective transformation β of $\mathbb{R}P^m$ is a projective map associated with a nonsingular matrix $B \in GL(m+1, \mathbb{R})$ and its action on β of $\mathbb{R}P^m$:

$$\beta([x^1 : \cdots : x^{m+1}]) = [B(x^1, \ldots, x^{m+1})] \tag{2}$$

- The projective transformation is given by $v = f(u)$, with

$$v^j = \frac{a^j_{m+1} + \sum_{i=1}^m a^j_i u^i}{a^{m+1}_{m+1} + \sum_{i=1}^m a^{m+1}_i u^i}, \forall j = 1, \ldots, m \tag{3}$$

where $\det B = \det((a^j_1)_{i,j=1,\ldots,m+1}) \neq 0$

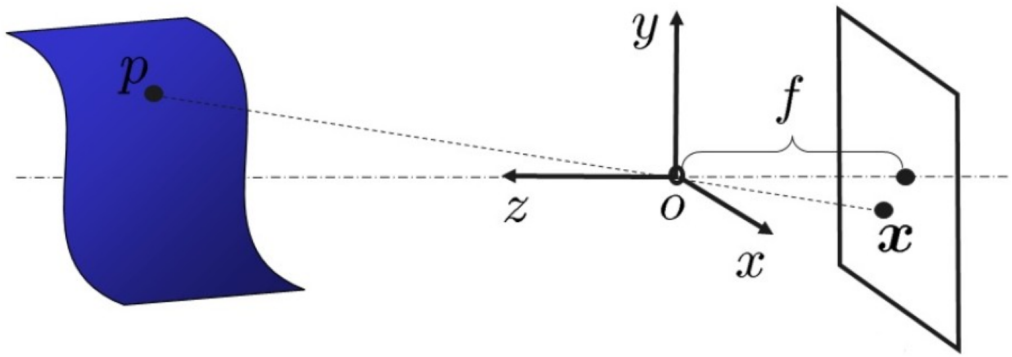
Projective Frame

A projective frame in $\mathbb{R}P^m$ is an ordered set of $m+2$ projective points in general position. An example of projective frame in $\mathbb{R}P^m$ is the standard projective frame $([e_1], \ldots, [e_{m+1}], [e_1 + \cdots + e_{m+1}])$.
In projective shape analysis it is preferable to employ coordinates invariant with respect to the group PGL(m) of projective transformations. A projective transformation takes a projective frame to a projective frame, and its action on $\mathbb{R}P^m$ is determined by its action on a projective frame, therefore if we define the projective coordinate(s) of a point $p \in \mathbb{R}P^m$ w.r.t. a projective frame $\pi = (p_1, \ldots, p_{m+2})$ as being given by

$$p^\pi = \beta^{-1}(p) \tag{4}$$

where $\beta \in \mathrm{PGL}(m)$ is a projective transformation taking the standard projective frame to π, these coordinates have automatically the invariance property.

Image Acquisition in Ideal Digital Camera



Ideal pinhole camera image acquisition can be thought of in terms of a central projection $\beta : \mathbb{R}P^3 \setminus \mathbb{R}P^2 \to \mathbb{R}P^2$, whose representation in conveniently selected affine coordinates $(x, y, z) \in \mathbb{R}^3$, $(u, v) \in \mathbb{R}^2$ is given by

$$u = f\frac{x}{z}$$
$$v = f\frac{y}{z}$$

where f is the focal length.

Image Acquisition in Ideal Digital Camera

- Consider the equation

$$\lambda \mathbf{x}' = K \Pi_0 g \mathbf{X} = K[R, T]\mathbf{X} \tag{5}$$

- $\lambda =$ depth of point
- $K = \begin{bmatrix} f \cdot s_x & s_\theta & o_x \\ 0 & f \cdot s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$: intrinsic parameter matrix
- $\Pi_0 = [I_3, 0_3]$
- $g \in \mathrm{SE}(3)$: pose of the camera in the reference frame

Image Acquisition in Ideal Digital Camera

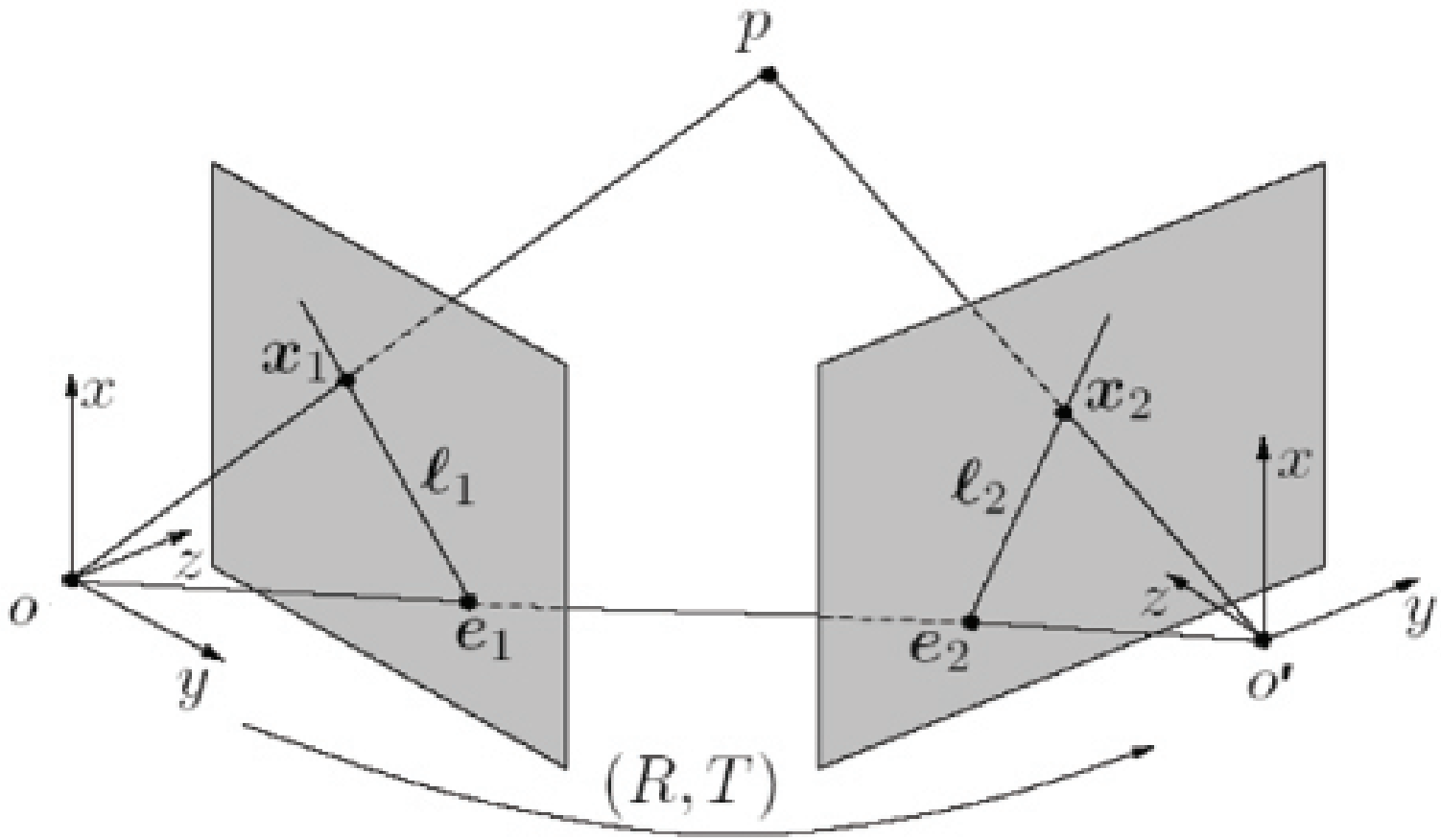


Figure: Two projections $\mathbf{x_1}, \mathbf{x_2} \in \mathbb{R}^3$ of a 3D point p from two vantage points.

Essential Matrix

- x_1 and x_2 are related by a rigid body transformation $(R, t) \in \mathrm{SO}(3) \times \mathbb{R}^3$

$$x_2 = Rx_1 + t \tag{6}$$

- x_2, $Rx_1 + t$ and t are direction of lines in the epipolar plane, we have

$$x_2^T (t \times (Rx_1)) = 0 \tag{7}$$

- By defining $t_\times$ as the matrix associated with the linear operation $y \to t \times y$, then

$$x_2^T (t_\times R) x_1 = x_2^T E x_1 = 0 \tag{8}$$

where E is so called essential matrix.

Fundamental Matrix

- If the camera is uncalibrated, then the matrices $K_1 = K_2 = K$ containing the camera initial parameters, yield the homogeneous pixel coordinates :

$$y_1 = Kx_1$$
$$y_2 = Kx_2$$

- Thus,

$$(K^{-1}y_2)^T E(K^{-1}y_1) = y_2^T (K^{-1})^T EK^{-1} y_1 = y_2^T F y_1 = 0 \tag{9}$$

where $F = (K^{-1})^T EK^{-1}$ is so called **fundamental matrix**.

Reconstruction

Problem The problem of the reconstruction of a configuration of points in 3D from two ideal uncalibrated camera images, is equivalent to the following: given two camera images $\mathbb{R}P_1^2$, $\mathbb{R}P_2^2$ of unknown relative position and unknown internal camera parameters, and two matching sets of labelled points $\{p_{a,1}, \ldots, p_{a,k}\} \subset \mathbb{R}P_a^2, \;\; a = 1, 2$ find a configuration of points $p_1, \ldots, p_k \in \mathbb{R}P^3$ such that there exist two positions of the camera $\mathbb{R}P_1^2$, $\mathbb{R}P_2^2$ for which $\beta_a(p_j) = p_{a,j}, \;\; \forall a = 1, 2, \;\; j = 1, \ldots, k$.

Theorem The reconstruction problem for two non calibrated camera images has a solution in terms of the fundamental matrix $F = t_\times R$. Any two solutions can be obtained from each other by a projective transformation in $\mathbb{R}P^3$.

Eight Point Algorithm

The linear system for F is

$$y_2^T F y_1 = 0 \tag{10}$$

and can be written as

$$Y^T f = 0 \tag{11}$$

where f is a vectorized form of F and Y is the tensor product of y_1 and y_2. Given a set of corresponding points, form the matrix of measurements $\chi \doteq [Y^1, Y^2, ..., Y^n]^T$; f can be obtained as the minimizing solution of the least-squares objective function $\|\chi f\|^2$. Such a solution corresponds to the eigenvector associated with the smallest eigenvalue of $\chi^T \chi$ and can be found using SVD.

Homogeneous spaces

Assume $\alpha : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ is a left action of $\mathcal{K}$ on $\mathcal{M}$ and define the **orbit** $\mathcal{K}(x)$ **of a point** $x \in M$ as the set $\{\alpha(k, x), k \in K\}$. Then $\mathcal{M}$ is a $\mathcal{K}$-**homogeneous space** if there is a point x s.t. $\mathcal{K}(x) = \mathcal{M}$. In the case of a **manifold** $\mathcal{M}$, we assume in addition that the action is smooth ($\mathcal{K}$ is a **Lie group**). In Statistics it makes sense to consider the equality of means, on an **smooth object space** $\mathcal{M}$, with an action of a Lie group $\mathcal{K}$, **only for means that on the same orbit**. We thus assume that the object space is a homogeneous space. Examples of **object spaces** are homogeneous spaces:

- spaces of directions ($\mathcal{M} = S^m, m = 1, 2$).
- spaces of dihedral angles ($\mathcal{M} = (\mathbb{S}^1)^k$).
- spaces of shapes of planar k-ad's ($\mathcal{M} = \mathbb{C}P^{k-2}$. (D.G. Kendall)
- spaces of shapes 2D coutours ($\mathcal{M} = (P(\mathbb{H})$, $\mathbb{H}$ Hilbert space)
- spaces of cell filaments ($\mathcal{M} = \mathbb{R}P^2 \times (0, \infty)$.

The Space of 3D Projective Shapes as a Lie Group

- A Lie group $(\mathcal{K}, \odot)$ act transitively on itself via left translations, providing a standard example of homogeneous space.
- Example: $\mathbb{R}P^3$ is a Lie group with the multiplication $[x] \odot [y] = [x \cdot y]$ for $[x], [y] \in \mathbb{R}P^3$. And the operation $(\cdot)$ is the quaternion multiplication (see Crane and Patrangenaru (2011)).
- In general $\mathbb{R}P^m$ is a homogeneous space. The group action $\alpha : SO(m) \times \mathbb{R}P^m \to \mathbb{R}P^m$ is given by

$$\alpha([A], [x]) = [Ax], \ \forall \ A \in SO(m), \ \forall \ x \in \mathbb{R}^{m+1} \backslash \{0\}$$

- It then follows that the 3D projective shape space of $k$ads $P\Sigma_3^k \sim (\mathbb{R}P^3)^q$ is also a homogeneous space, at it inherits a product Lie group structure from $\mathbb{R}P^3$.

This homogeneous space is of particular interest in **3D vision**

VW mean of random 3D projective shapes

The projective shape space of k-ads $P\Sigma_3^k = (\mathbb{R}P^3)^{k-5}$ is embedded in the linear space $(Sym(4,\mathbb{R}))^{k-5}$ via the so called **VW embedding** $j_q, q = k-5$.
Let Q a **(V-W) nonfocal probability measure** on $(\mathbb{R}P^3)^{k-5}$, $q = k-5$, with

$$Y = ([X^1], ..., [X^q]),$$
$$(X^s)^T X^s = 1, \quad \forall\, s = 1, \ldots, q$$

where $[X^s] \in \mathbb{R}P^3$.
The **V-W mean** on $(\mathbb{R}P^3)^q$ is given by;

$$\mu_{j_k} = (\gamma_1(4), \ldots, \gamma_q(4)) \tag{12}$$

where $(\lambda_s(a), \gamma_s(a))$, $a = 1, 2, 3, 4$ are the eigenvalues in increasing order and the corresponding unit eigenvectors of the matrix $E[X^s(X^s)^T]$.

Two sample hypothesis test for means on a Lie group

In the case of a Lie group $\mathcal{M} = \mathcal{K}$, one two sample hypothesis testing problem for extrinsic means on the embedded manifold $j : \mathcal{M} \to \mathbb{R}^N$, given a fixed element $\delta \in \mathcal{K}$, one may consider the two problem of testing for the extrinsic means of two random objects the following

$$H_0 : \mu_{2,j} = \delta \cdot \mu_{1,j}, \quad \textit{versus} \quad H_1 : \mu_{2,j} \neq \delta \cdot \mu_{1,j}$$

This is equivalent to the following hypothesis testing problem,

$$H_0 : \mu_{2,j}) \odot \mu_{1,j}^{-1} = \delta, \quad \textit{versus} \quad H_1 : \mu_{2,j} \odot \mu_{1,j}^{-1} \neq \delta$$

We now let $H : \mathcal{M}^2 \to \mathcal{K}$, defined by $H(x_1, x_2) = x_2) \odot x_1^{-1}$ is smooth, therefore large sample methods on manifolds, we may be used to derive the asymptotic behavior of the $H(\bar{X}_1, \bar{X}_2)$ where $\bar{X}_a$ is the extrinsic sample mean for the a-th sample, $a = 1, 2$. Here we consider unmatched pairs on $\mathcal{K}$.

Bootstrap distribution on $\mathcal{M} = (\mathbb{R}P^3)^{k-5}$

For $a = 1, 2$, let $\{Y_{a,r}\}_{r=1}^{n_a}$ be i.i.d. random samples defined on $\mathcal{M}$ from independent Veronese-Whitney-nonfocal distributions $\mathcal{Q}_a$.

•Assume $n = n_1 + n_2$ and $\lim_{n\to\infty} \frac{n_1}{n} \to \pi \in (0, 1)$. Let φ and L_δ be respectively, the log chart and the left translation by $1_{\mathcal{M}} \in \mathcal{M}$. Then under H_0, the joint distribution of

$$D = \left(\varphi \circ L_\delta^1(H(\bar{Y}_{n_1,j}, \bar{Y}_{n_2,j}))\right) = \varphi(\bar{Y}_{n_2,j_k} \otimes \bar{Y}_{n_1,j_k}^{-1})$$

can be approximated by the bootstrap joint distribution of

$$D^\star = \varphi(\bar{Y^\star}_{n_2,j_k} \otimes \bar{Y^\star}_{n_1,j_k}^{-1})$$

with an error $O_p(n^{-1/2})$.

•We can construct a nonparametric bootstrap c.r. of size $1 - \alpha$, for simplicity by using simultaneous confidence intervals. We will then expect to **fail to reject** the null, if we have 0 in all of our simultaneous confidence intervals for the affine coordinates.

Data set and landmark placement

The data set analyzed consists in pictures of faces:

Figure: Faces Used for Analysis

For analysis, 10 landmarks were placed, resulting 3D projective shapes of k-ads ($k = 10$).

Data set and landmark placement

The collections and reconstructions of all of our landmark configurations were done in MATLAB. The landmarks are shown below;

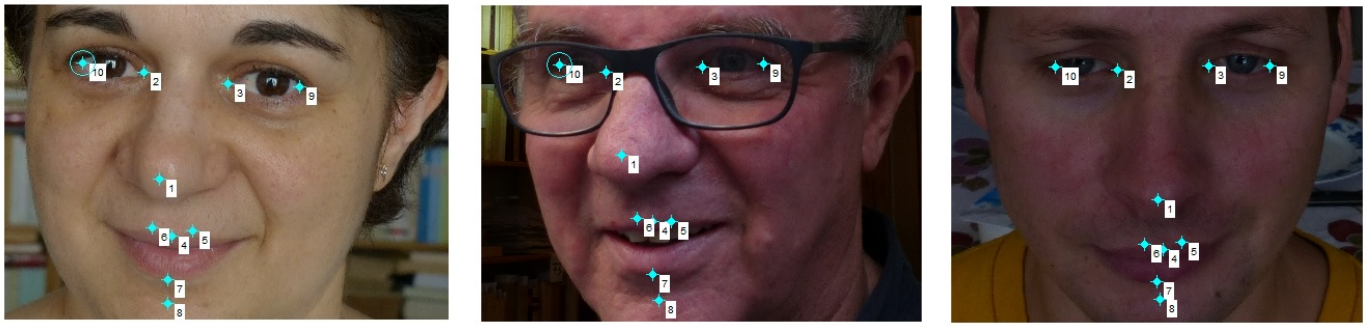


Figure: Landmark Placements for all faces

The first five landmarks were used as projective frame, which are, in increasing order: *pronasale, right and left Endocathion, Labiale Superius, left Crista Philtri*.

Comparing male faces

For our result we used simultaneous confidence intervals mentioned above. We failed to reject the null hypothesis

$$H_0 : \mu_{1,10}^{-1} \odot \mu_{2,10} = 1_5$$

if all of our confidence intervals contain the value 0. For this two sample test, we use the following two sets of data;

Figure: Male Face Sample 1

Bootstrap marginals and simultaneous confidence intervals for male faces

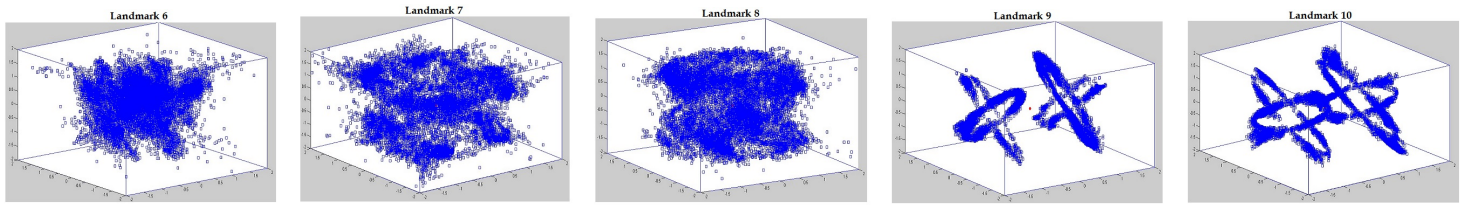


Figure: Bootstrap projective shape marginals for male face data

Simultaneous confidence intervals for 2 sample face's landmarks 6 to 8			
	LM6	LM7	LM8
x	$(-1.111498, 0.805386)$	$(-1.117512, 1.099536)$	$(-1.296547, 0.966296)$
y	$(-1.215218, 0.710931)$	$(-1.355167, 1.336021)$	$(-0.635282, 1.372627)$
z	$(-1.161234, 1.150762)$	$(-1.432217, 1.349541)$	$(-1.394141, 1.349442)$

Simultaneous confidence intervals for 2 sample face's landmarks 9 and 10		
	LM9	LM10
x	$(0.952164, 0.996354)$	$(-0.962541, 1.005917)$
y	$(-0.760124, 1.129782)$	$(-1.070631, 0.982195)$
z	$(-0.817503, 1.319117)$	$(-1.319374, 1.089272)$

We notice that one of the simultaneous confidence intervals for landmark 9, corresponding to the right *Exocanthion*, does not contain 0. We then reject the null hypothesis, showing that there is significant projective shape change between the two male faces.

Comparing female faces

To get the simultaneous confidence intervals for our null hypothesis

$$H_0 : \mu_{1,10}^{-1} \odot \mu_{2,10} = 1_5,$$

we use the picture below as our data;

Figure: Female 1 Sample

Figure: Male 2 Sample

Bootstrap marginals and simultaneous confidence intervals for cross gender analysis

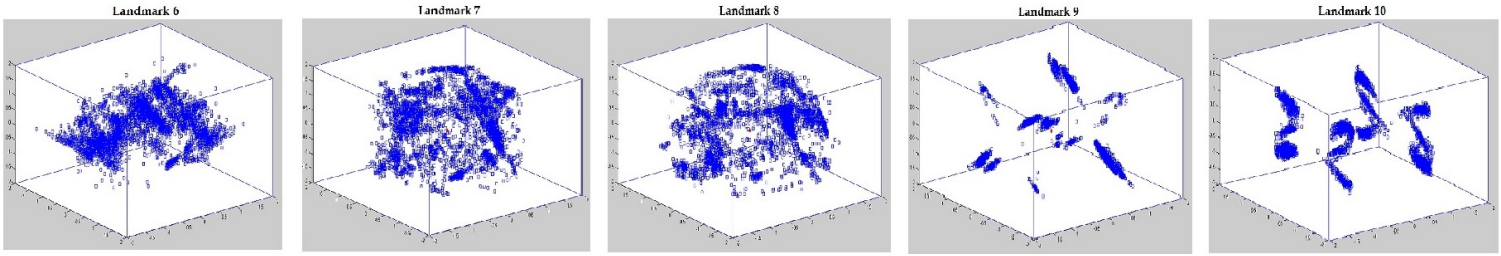


Figure: Bootstrap projective shape marginals for cross gender data

Simultaneous confidence intervals for cross gender for landmarks 6 to 8			
	LM6	LM7	LM8
x	$(-1.251984, 1.202986)$	$(-1.228628, 1.234229)$	$(-1.273092, 1.332798)$
y	$(-0.633834, 0.902621)$	$(-0.928523, 0.995304)$	$(-0.226587, 0.865510)$
z	$(-0.231190, 0.432009)$	$(-0.684483, 1.045302)$	$(-0.590623, 1.132418)$

Simultaneous confidence intervals for cross gender for landmarks 9 and 10		
	LM9	LM10
x	$(0.998446, 1.028374)$	$(-0.988191, -0.931250)$
y	$(-0.702335, 0.540613)$	$(-1.162803, 1.008259)$
z	$(-1.057821, 0.849069)$	$(-0.118635, 0.969739)$

The landmark 9 and 10 corresponding to the right and left Exocanthion have intervals not containing 0. We reject the null hypothesis, and conclude that there is a significant projective shape change between the two faces.

Cross validation

We separate the original sample into two smaller data sets of sizes $n_1 = 5$ and $n_2 = 6$. They are displayed below;

Figure: Cross Validation Sample 1

Figure: Cross Validation Sample 2

Bootstrap marginals and simultaneous confidence intervals for cross validation

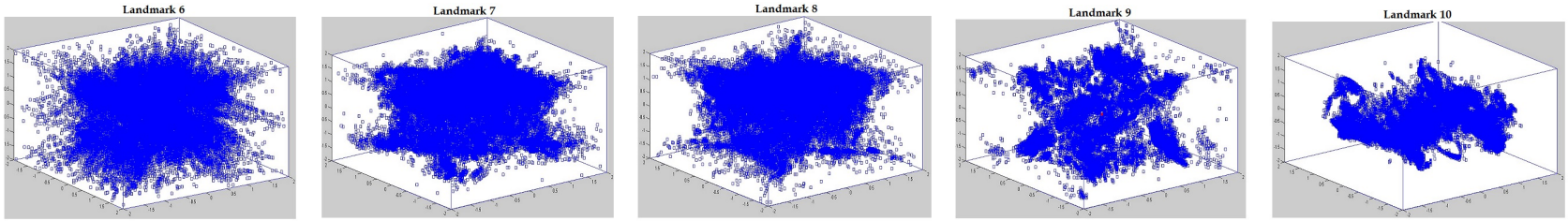


Figure: Bootstrap Projective Shape Marginals Cross Male Face 2

Simultaneous confidence interval for cross validation face 2 for landmarks 6 to 8			
	LM6	LM7	LM8
x	$(-17.496785, 3.552070)$	$(-4.027879, 4.860970)$	$(-1.990796, 7.497709)$
y	$(-10.967285, 4.340129)$	$(-3.776026, 9.830274)$	$(-7.558584, 0.865119)$
z	$(-2.724184, 13.093615)$	$(-3.006049, 5.891478)$	$(-0.698745, 4.293201)$

Simultaneous confidence intervals for cross validation face 2 for landmarks 9 and 10		
	LM9	LM10
x	$(-2.459882, 1.230096)$	$(-3.264292, 1.036499)$
y	$(-1.631839, 0.983147)$	$(-1.387133, 2.942318)$
z	$(-1.451487, 1.196335)$	$(-0.916768, 1.658124)$

All the simultaneous intervals contain 0 and thus we fail to reject the null hypothesis. And we report no statistically significant mean projective shape change.

Discussion

- Face Identification
  - Projective Frame
  - 3D reconstruction as a model
- Illumination will affect detection of corresponding features (landmarks).
- For this reason an analysis of color scenes is useful (see Patrangenaru and Deng(2020))
- Colored scenes under identical illumination help automatic landmark correspondence detection

Reference

- Ma, Yi, et al. An invitation to 3-d vision: from images to geometric models. Vol. 26. New York: springer, 2004.
- Patrangenaru, Victor, and Leif Ellingson. Nonparametric statistics on manifolds and their applications to object data analysis. CRC Press, Taylor & Francis Group, 2016.
- Jain, Anil K., and Stan Z. Li. Handbook of face recognition. Vol. 1. New York: Springer, 2011.
- V. Patrangenaru, X. Liu and S. Sugathadasa (2010). Nonparametric 3D Projective Shape Estimation from Pairs of 2D Images - I, In Memory of W.P. Dayawansa. *Journal of Multivariate Analysis*. **101**, 11-31.
- V. Patrangenaru, Y. Deng (2020). Nonparametric Data Analysis on the Space of Perceived Colors. *arXiv:2004.03402*