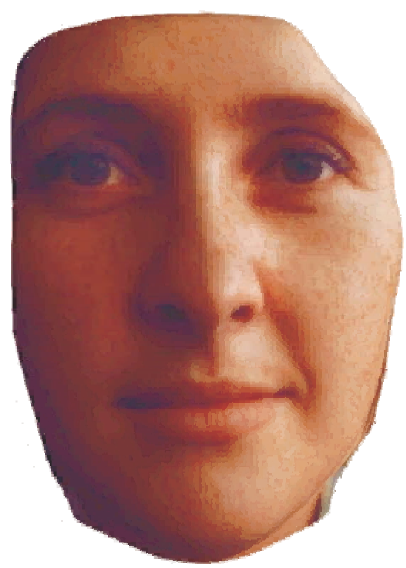
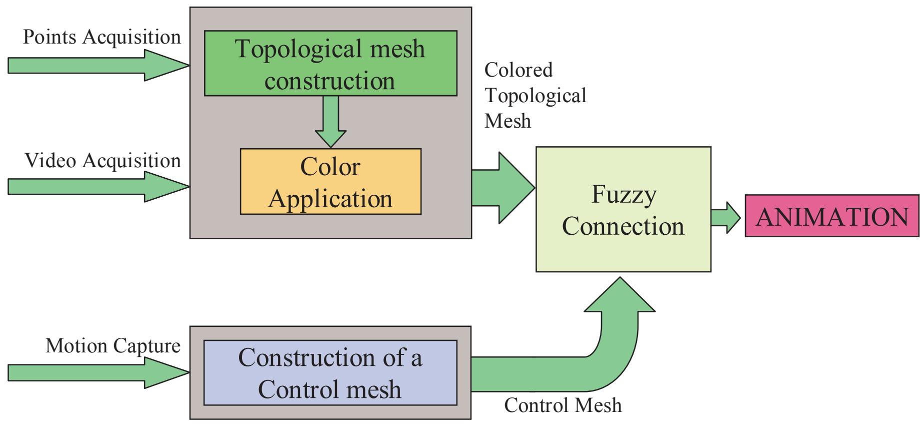
From Real faces to Virtual faces

Alberto Borghese
Department of Computer Science
University of Milano

Which is real, which is virtual?



General schema



Laboratory of Motion Analysis & Virtual Reality, MAVR

http://homes.dsi.unimi.it/~borghese

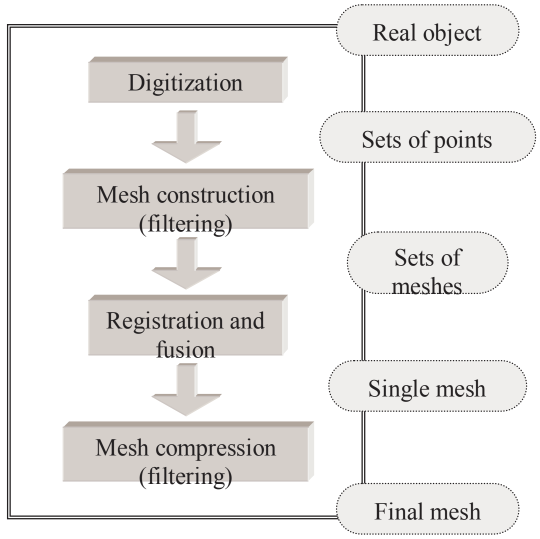
Outline

- Points acquisition
- From points to surface (mesh)
- Mesh compression
- Application of colour attribute
- Animation

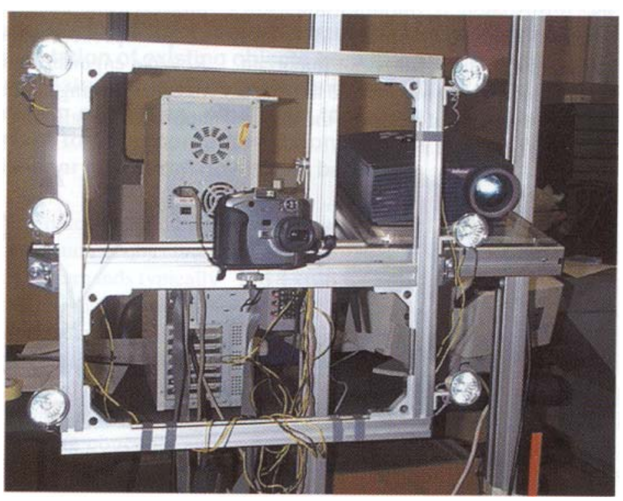
Laboratory of Motion Analysis & Virtual Reality, MAVR

http://homes.dsi.unimi.it/~borghese

Construction of the topological mesh



In-house Digitizers



- Projection of patterns through a standard video projector.
- Imaging through standard photocameras.
- Image processing to extract range data points and texture.

•There is some rigidity in the distribution of the range data.

Minolta digitizers

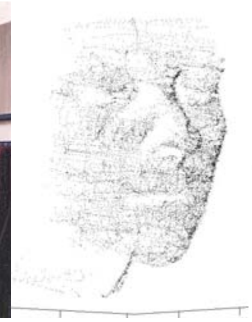
- **Speed** - scans in less than one second (Fast Mode)
- **Precision** - over 300,000 points with range resolution to 0.0016" (Fine Mode)
- **Simplicity** - point and shoot simplicity for consistently excellent results
- **Flexibility** - only Minolta offers interchangeable lenses for variable scanning volumes

http://www.minolta-3d.com/

Face digitization (Autoscan)

Acquisition session

Points cloud

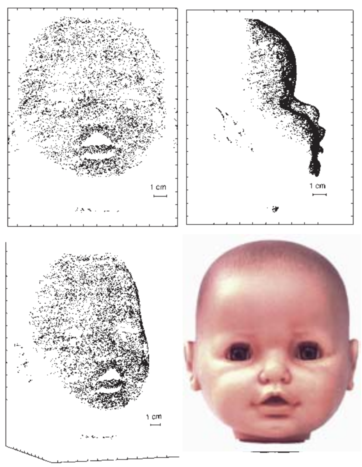


Direct tessellation

- Pair of video-cameras + standard laser pointer.
- The range data are obtained by "painting" the surface manually.
- Set of range data, which is denser where required.
- High precision in spot localization (cross-correlation, bright image).

Drawback: High scanning time. ***Direct tessellation produces an undesirable result.***

Digitization introduces errors



Interpolation schemes (e.g. Delauney tessellation) fails because of measurement noise.

⇒

The need of filtering is evident.

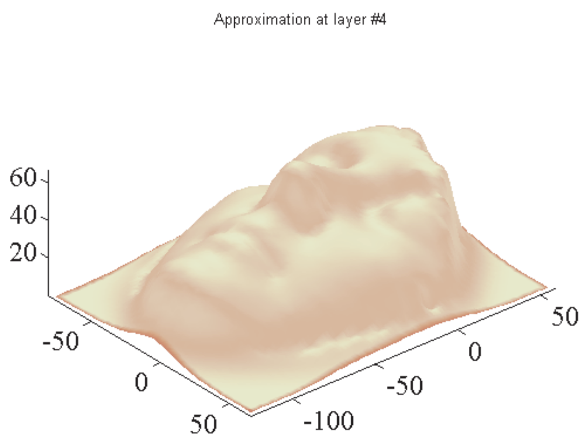
How to convert the points into a mesh?

Problem: noise
Solution: regularized solutions.

• Human body parts are “smooth” (lisce).
• Noise has spatial frequencies higher that surface.
• Surface has been over-sampled.

HRBF Networks

Incremental Reconstruction, error-driven.



Gaussian RBF Networks

Linear combination of Gaussian functions: $\mathbf{z} = s(\mathbf{x}) = \sum_{k=1}^{M} \mathbf{w}_k G(\mathbf{x}; \mathbf{c}_k, \Sigma_k)$

Pioneers in exploring properties of quasi-local units:
- Broomhead e Lowe, 1988.
- Moody e Darken, 1989.
- Poggio e Girosi, 1990.
- Park e Sandberg, 1991.

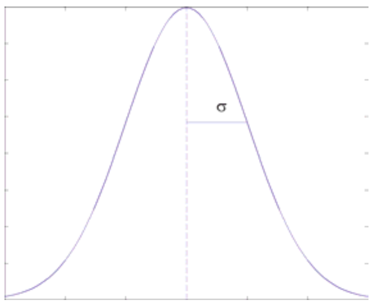
Linear filtering:
- Sanner and Slotine, 1992.
- Canon and Slotine, 1995.
- Borghese and Ferrari, 1996, 2001.
- Poggio et al., 1993.
- Canny, 1986 (Computer vision domain).

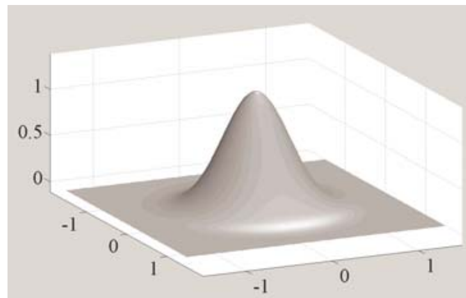
Gaussian functions

$$G(P;\mu\,|\,\sigma) = \frac{1}{\pi^{D/2}\sigma^{D}} e^{-\frac{\|P-\mu\|^2}{\sigma^2}}$$

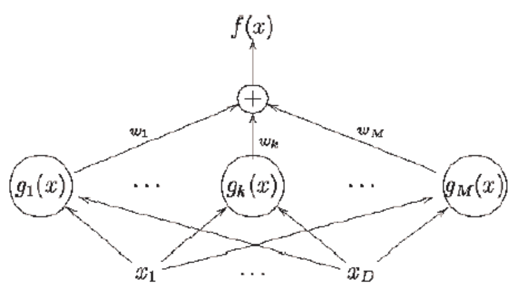
$P, \mu \in R^D$

$\sigma \in R$





The RBF parameters



$$g_k(x) = \frac{1}{\sqrt{\pi^D}\sigma_k^D} exp\left(-\frac{\|x-\mu_k\|^2}{\sigma_k^2}\right)$$

$x, \mu_k \in \mathcal{R}^{\mathcal{D}}, \sigma_k \in \mathcal{R}$

M, μ_k, σ_k are the structural parameters.
w_j are the synaptic weights.

Learning strategies

•**Empirical models:** *Broomhead and Lowe, 1988; Moody and Darken, 1989; Park and Sandberg, 1991.*

•**Regularisation theory:** *Yuille e Grzywacz, 1988; Poggio and Girosi, 1990; Girosi et al., 1995; Wahba and Xu, 1998.*

The parameters: M, $\{P_k\}$ and σ_k are set through an optimization process. Sparse approximation *but* non linear optimization-

•**Filtering Theory:** *Sanner e Slotine, 1992; Canon and Slotine 1995, Borghese and Ferrari, 1996, 2001; Canny, 1986.*

Linear Gaussian filter

$$\mathbf{z} = s(\mathbf{x}) = \sum_{k=1}^{M} \mathbf{w}_k G(\mathbf{x};\mathbf{c}_k,\Sigma_k)$$

Let us suppose: $\Sigma k=\Sigma\ \forall k \quad (c_{k+1} - c_k) \to 0$

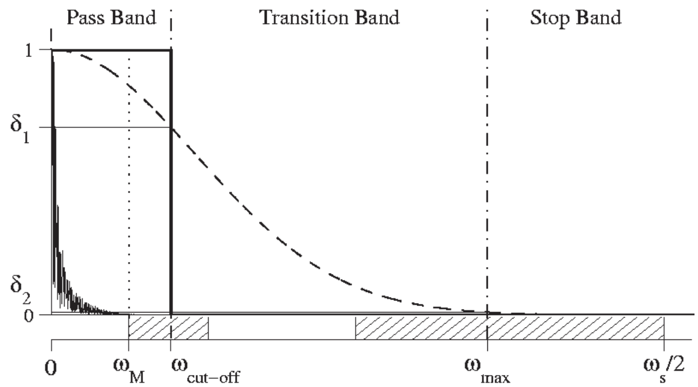
Continuos RBF: $s(x) = \int_R w(c) G((x-c)|\sigma) dc$

STATEMENT 1: Let w(x), s(x) and $G(x-c|\sigma) \in L1(R)$ and be invariant to translation, then the continous **RBF Network is equivalent to the *convolution*** of the function w(x) with the Gaussian function: $s(x) = w(x)*g(x;\sigma)$

In the frequency domain: $S(\nu) = W(\nu)\, G(\nu;\sigma)$

$W(\nu)$ plays the role of a noisy version of $S(\nu)$.

Low-pass behavior of the Gaussian filter



Pass band [0 $\nu_{cut\text{-}off}$]

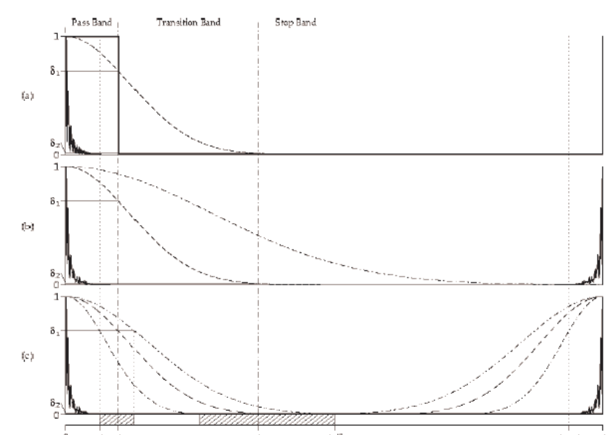
Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese

Discrete linear filter

$c_{k+1} - c_k = \Delta c \ \forall c_k$
Equally spaced Gaussians

$$\mathbf{z} = s(\mathbf{x}) = \sum_{k=1}^{M} \mathbf{w}_k G(\mathbf{x}; \mathbf{c}_k, \Sigma_k)$$

Output: interpolation through the Gaussian basis

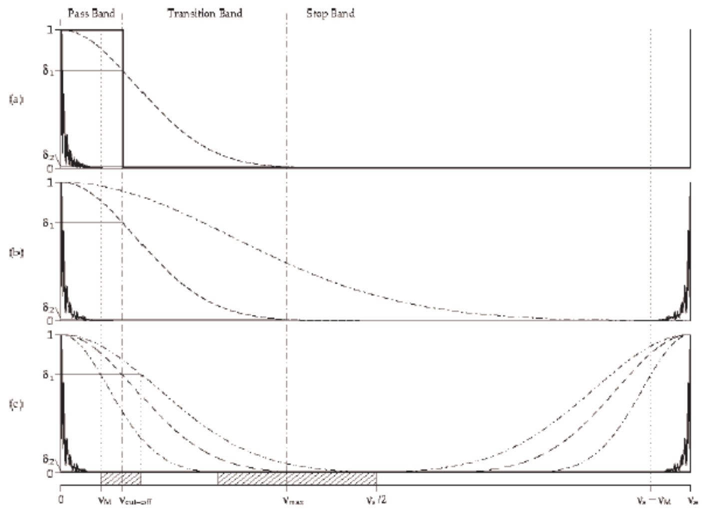


Pass band [0 $\nu_{cut\text{-}off}$]
Stop band [ν_M $\nu_{s/2}$]

$\nu_s = 1/\Delta c$

Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese

Low-pass behavior of the Gaussian filter



Pass band $[0\ \nu_{cut\text{-}off}]$ Stop band $[\nu_M\ \nu_{s/2}]$

Gaussian filtering

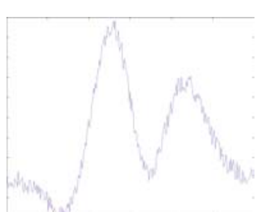
Linear combination of gaussians

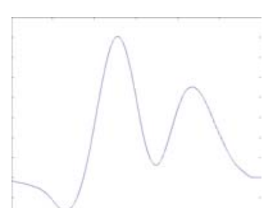
$$s(x) = \sum_{k=1}^{M} w_k\, g(x; x_k, \sigma_k)$$

single σ + regular spacing

low-pass gaussian filter

$$s(x) = w_k * g(x - x_k; \sigma)$$





$$S_l(P) = \sum_{k=1}^{M_l} z(P_{kl})\, G(P; P_{kl} \mid \sigma_l)\, \Delta P_l^2$$

Linear filtering

Artificial Vision. Filter grids which operate at different scales.

Small scale (high frequency)

Large scale (low frequency)

Linear combination of Basis Functions: $S(P) = \sum_k^M S_k G\left(P - P_k \mid \sigma\right)$

HRBF Networks operation

Quasi-local operations => **Receptive field**.

•$S(P_k)$ is estimated through a local weighted mean in the grid crossings:

$$S(P_k) = \frac{\sum_m S_m(x_m) G(x_m - x_k \mid \sigma)}{\sum_m G(x_m - x_k \mid \sigma)}$$

• x_m belongs to the neighbourhood of $G(x_m-.|.)$

$$s(x) = \sum_{k=1}^{N} S(P_k) G((x - x_k); \sigma)$$

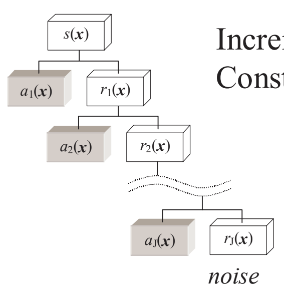
Problem with this approach

Single scale. 3D objects have different scales in different spatial locations.

Small scale. There may be not enough points inside the receptive field of a Gaussian function. Small scale $\Rightarrow$ Dense packing.

Solutions:
•Wavelets (from fine to coarse).
•Adaptive hierarchical approach.

HRBF Networks



Incremental
Construction of the network

$$\frac{\sum_{P_r \in A(P_{kl})} |r_1(P_r)|}{N_k} < \varepsilon_n \quad \approx \text{noise}$$

HRBF Networks operation

Quasi-local operations => **Receptive field**.

•S(P_k) is estimated through a local weighted mean in the grid crossings:

$$S(P_k) = \frac{\sum_m S_m(x_m) G(x_m - x_k \mid \sigma)}{\sum_m G(x_m - x_k \mid \sigma)}$$

•A residual is computed for each measured point m as:

$$R(P_m) = S(P_m) - S_m(P_m)$$

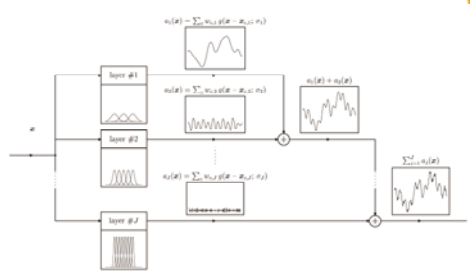
• The local reconstruction error is evaluated with a local integral metric for each crossing k as:

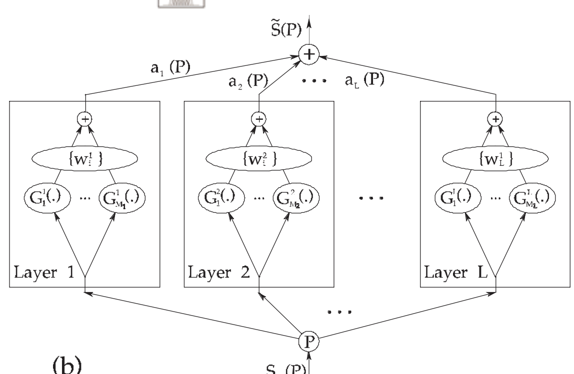
$$MRE(P_k) = \frac{\sum_m \mid R_m(x_m) - S(x_m) \mid}{N_k}$$

Hierarchical Radial Basis Function Network

The surface is therefore reconstructed as:

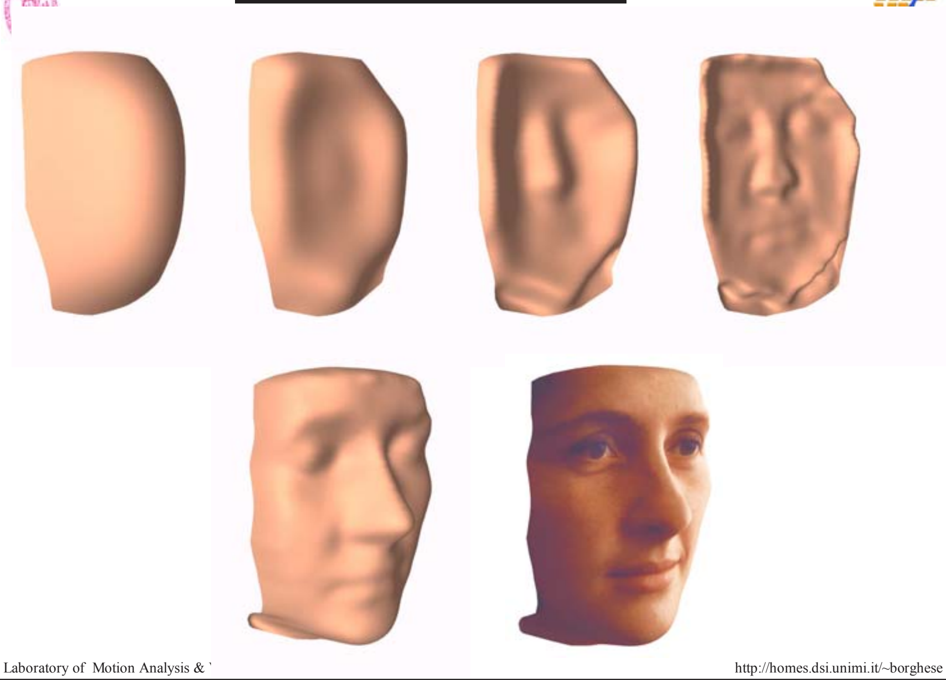
$$S(P) = \sum_{l=1}^{L} \sum_{k=1}^{M_l} w_{kl} G((P - P_{kl}) \mid \sigma_l)$$



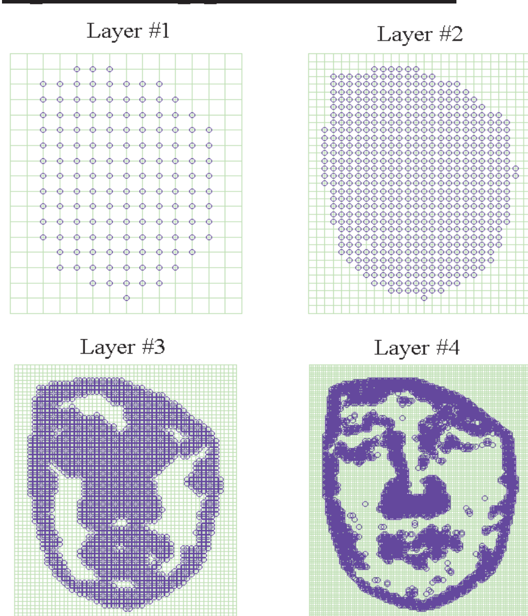


(b)

Incremental Reconstruction



Sparse approximation



Setting the parameters of the HRBF

From linear filtering theory: $S_r(P) = \sum_{l=1}^{L} a_l(P) = \sum_{l=1}^{L} \sum_{k=1}^{M_j} a(P_{kl}) G(P - P_{kl} \mid \sigma_l) \Delta\mu_l^2$

First layer: $a(P_{k1})$ estimates $S(P_k)$

l-th layer: $a(P_{kl})$ estimates $r_{l-1}(P_{kl}) = S_m(P_l) - \sum_{j=1}^{l-1} \sum_{k=1}^{M_j} a(P_{kj}) G(P_l - P_{kj} \mid \sigma_j) \Delta\mu_j^2$

m = Measured (sampled point) $\sigma_l = \sigma_{l-1}/2$

r = Reconstructed through HRBF $\Delta\mu_l = \Delta\mu_{l-1}/2$

- $a(P_{kl})$ is determined through a local *Maximum a-posteriori estimate*:

$$a(P_{kl}) = \frac{\sum_{h=1}^{R_{kl}} a(P_h) G(P_h - P_{kl} \mid \sigma_l)}{\sum_{h=1}^{R_{kl}} G(P_h - P_{kl} \mid \sigma_l)}$$

$P_h \in$ Receptive field of $G(P_{kl} \mid \sigma_l)$

- A Gaussian $G(P_{kl})$ is inserted in the grid only if: $\frac{\sum_{h=1}^{R_{kl}} \left| r_{l-1}(P_h) \right|}{R_{kl}} < \varepsilon$ (noise)

Hierarchical Radial Basis Functions Network (Borghese and Ferrari, 1998 - Neurocomputing)

- Stacking grids of Gaussians one over the other
- Computation of the parameters with local operations (no “learning” = no iterations)
- Uniform reconstruction error

It belongs to the family of
“Incremental Surface-Oriented Reconstruction”
(Mencl and Muller, 1998)

HRBF summary

•Stacking grids (not complete) of Gaussians one over the other. Sparse approximation.

•Quasi-local operations => **Receptive field**.
•High parallelism.

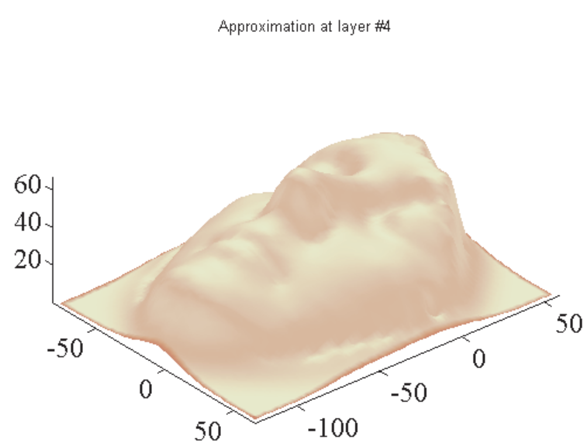
•Computation of the parameters with local operations (no “learning” = no iterations).

•Uniform residual error ≈ measurement noise.

It belongs to the family of “Incremental Surface-Oriented Reconstruction” (Mencl and Muller, 1998).

HRBF Networks

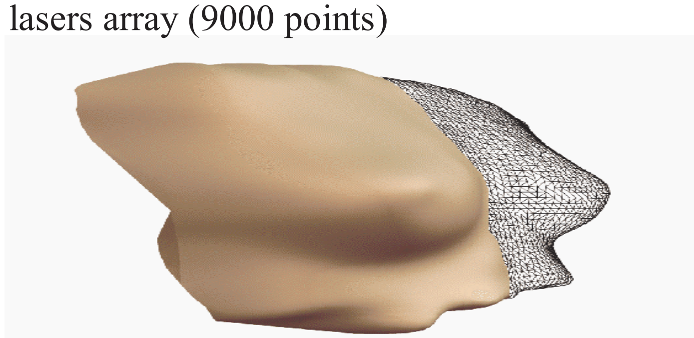
Incremental Reconstruction, error-driven takes a few seconds



Surface sampling and mesh semplification

Results on breast reconstruction

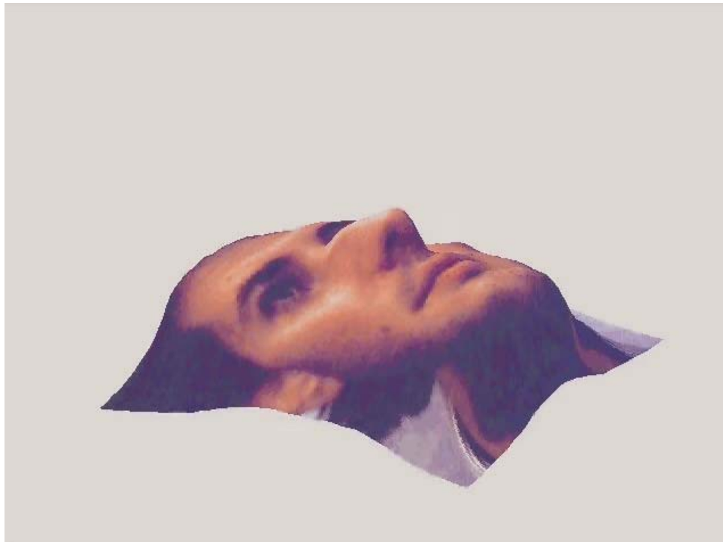
• 7 subjects
• 30 secs aquisition
• 3 lasers array (9000 points)



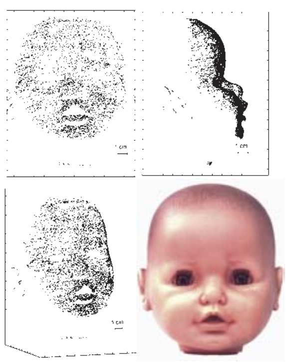
Error estimation:

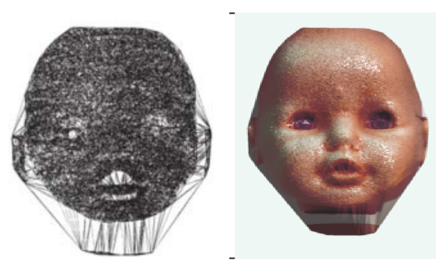
• Volume computation error: 4.3 %
• Linear error: 2 %
• Surface topography: 5 %

Results on morphing



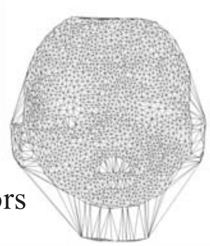
Data compression





Interpolation schemes (e.g. Delauney tessellation) fails, because of measurement noise.

100,000 data points



2,000 Reference Vectors

Mesh compression. Why?

•Limited bandwidth, limited capacity of processing and memory.
•Simplification of mesh processing.

•Compression - Transmission – Decompression.

•**Two large families**: *lossy or non-lossy compression.*
•
•• Lossy compression. The information lost in not relevant to the data usage. For example, here we want to loose noise.

Vector Quantization (VQ)

•The data are approximated with a reduced data set of points called *reference vectors*.

Given:
-a set $V \subseteq R^n$ of N data points ($v_1, v_2 \dots v_N$).
-a set $W \subseteq R^n$ of M reference vectors ($w_1, w_2 \dots w_M$).

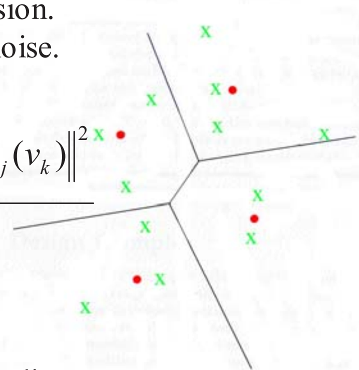
The W are a vector quantization coding of V if a certain function of V and W is minimized.

We define as winning reference vector: $\min_{w_j}\left(\|v - w_j\|^2\right)$

Compression through VQ

• Techniques widely used for lossy compression.
• They are used here to **loose** the digitizing noise.

$$E(V,W)=\frac{\sum_{k=1}^{N} d\left(v_k, w_j(v_k)\right)^2}{N}=\frac{\sum_{k=1}^{N} \left\| v_k - w_j(v_k) \right\|^2}{N}$$



• V is input: the set of range data of cardinality N.
• W is the output: a reduced set of points, of cardinality M << N called Reference Vectors.

Hard learning

$$E=\int_V P(v)\left(v-w_{i(v)}\right)^2 dv$$

$$\Delta w_i = \varepsilon \delta_{i(v)} (v - w_i)$$

Displacement of only 1 RV for each data point.

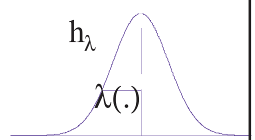
Soft-Max adaptation

•For each iteration, t, extract a data point $\tilde{v}(t)$

•For all the Reference Vectors, w, compute a displacement, such that E(V,W) decreases:

$$\Delta w_j(t) = \varepsilon(t) \cdot h_{\lambda(t)}\left(k_j\left(\tilde{v}(t), w_j\right)\right)\left(\tilde{v}(t) - w_j\right)$$

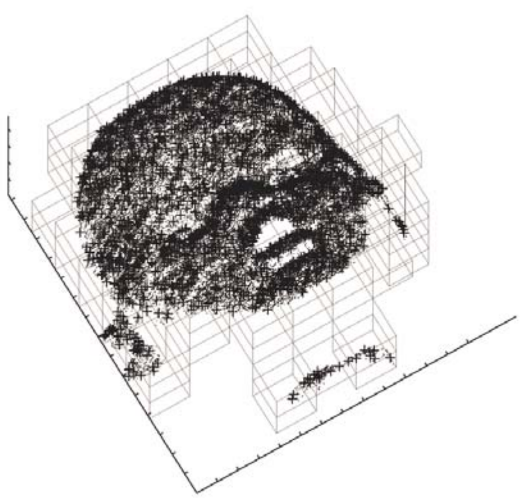
$h_{\lambda(t)}$ determines the receptive field for $\tilde{v}(t)$
Its amplitude decreases as optimization progresses.



• The w_js receive an adaptation, which decreases with their distance from $\tilde{v}(t)$

• Good solution in a computational time O(NM logM).

Speed-up through HB Processing



Enhanced Vector Quantization (EVQ).

How to use the HBs?

• "Intelligent" initialization. RVs are distributed such that their asymptotical density is locally observed:

$$M_k = M \frac{{N_k}^\gamma}{\sum_{k=1}^{N_H} {N_k}^\gamma} = \tau N \frac{{N_k}^\gamma}{\sum_{k=1}^{N_H} {N_k}^\gamma}$$

γ depends on the space dimension (Zador, 1982): γ = D/(D+2)

M_k does not depend on metric information.

Optimization

$$\Delta w_j(t) = \varepsilon(t) \cdot h_{\lambda(t)}\left(k_j\left(\tilde{v}(t), w_j\right)\right)\left(\tilde{v}(t) - w_j\right)$$

A receptive field is defined for $\tilde{v}(t)$

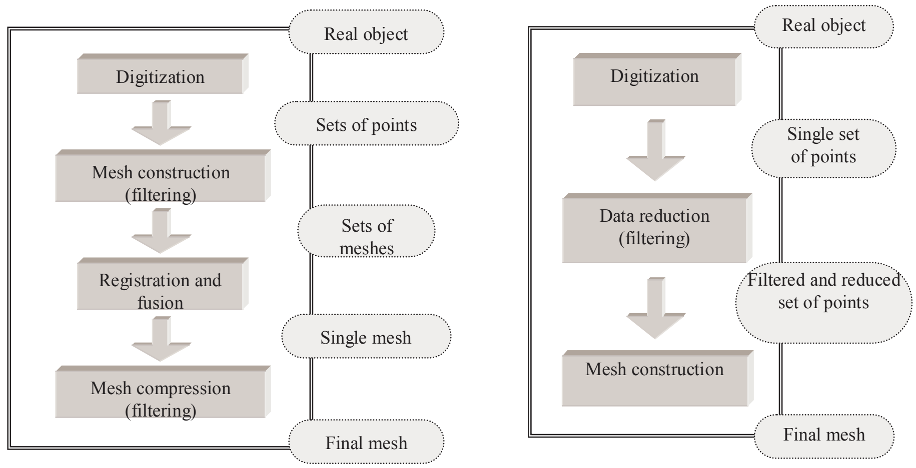
Ordering and updating is only for those w_j inside the receptive field.

•The receptive field is defined here as the 2^D closest boxes.
w_js far from $\tilde{v}(t)$ receive little updating because the Receptive Field shrinks.

Parameters setting

•Algorithms for automatically setting the parameters: ε, λ, the box side, have been derived.

New pipe-line for mesh construction

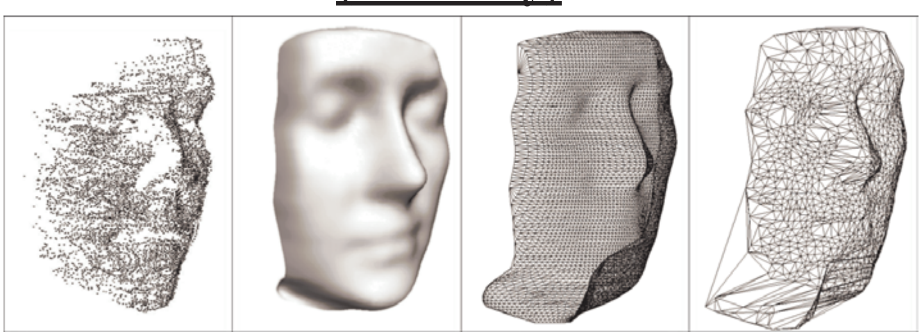


Traditional pipe-line

Pipe-line proposed here.

Mesh is constructed and managed from a reduced set of (filtered) points

Constructiong the topological mesh (summary)



Hypothesis: surfaces are smooth.

- Regularized reconstruction through HRBF networks.
- Regular dense sampling.
- Mesh compression through VQ.

Result: a topological mesh, geometrically accurate.

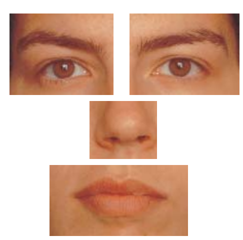
Range data from Raw Video sequences

- Camera internally calibrated (metric cameras).
- At least two pictures.

- Automatic identification of a set of features to initialize Bundle Adjustment or estimate the Essential matrix and identify additional corresponding points through them.

Identification of sub-images

1. Three regions clustering.
2. Vertical & horizontal projection of middle range region.
3. Identification of the peaks in horizontal and vertical histograms and geometrical considerations leads to the identification of sub-regions.



Features found

- Sobel gradient detector
- Colour clustering.
- Curve fitting.

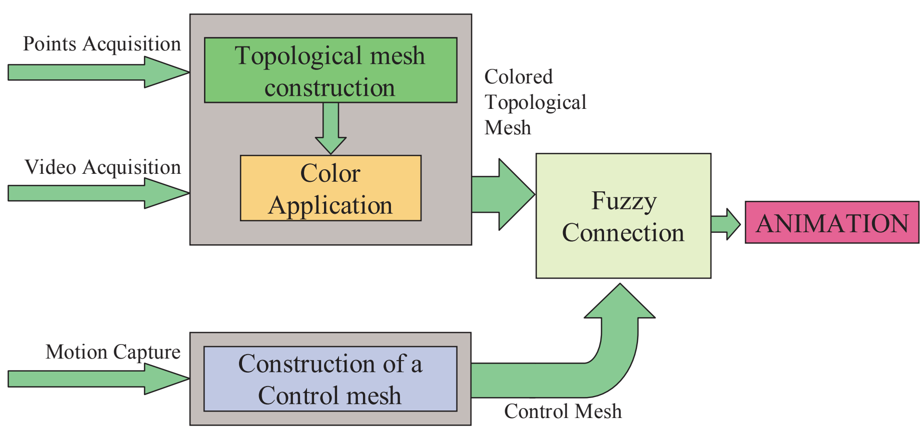
•Algorithm tested on a database of 46 images.

• Processing time 14s to 2 minutes with interpreted language (IDL) which rises hopes for real-time.

Laboratory of Motion Analysis & Virtual Reality, MAVR

http://homes.dsi.unimi.it/~borghese

General schema

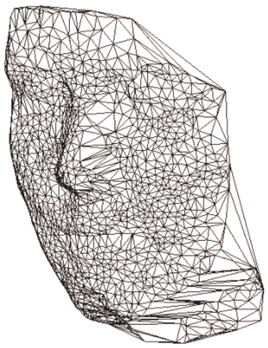


Laboratory of Motion Analysis & Virtual Reality, MAVR

http://homes.dsi.unimi.it/~borghese

Colour Application

Topological mesh | Bitmap | Coloured Mesh



Gouraud shading: vertex colour interpolation => colour field.
Low power graphics and soft shading.

Colour

Colour is the colour which is perceived, seen, that is the colour which is reflected by the objects surface.

Grey level images: white -> black.

Colour images: Red Green Blue (additive mix). Same primary colours present in the human retina.

Hue. Describes the colour (red, green…)
Saturation. Quantity of the colour. It differentiates red from rose. It can be viewed as the difference from the colour and a grey with the same brightness.
Lightness. Intensità del colore, it depends on the hue and saturation. It can be viewed as the colour of the image in B/W. It is due to the illumination intensity.

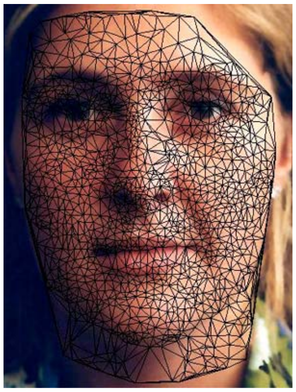
Colors (examples)

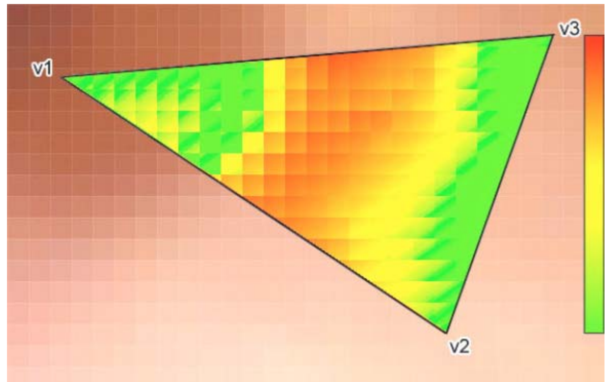
☐	White	(R=255, G=255, B=255)
■	Grey	(R=200, G=200, B=200)
■	Dark grey	(R=200, G=200, B=200)
■	Black	(R=0, G=0, B=0)
■	Red	(R=255, G=0, B=0)
■	Yellow	(R=255, G=255, B=0)
■	Pale blue	(R=0, G=255, B=255)
■	Green	(R=0, G=200, B=0)

Picking up the vertexes colour

•Alignment: porjection of topological mesh onto the bitmap plane.
•Colouring: Assigning the colour of the proper pixel to the vertexes.

Chromatic scanning



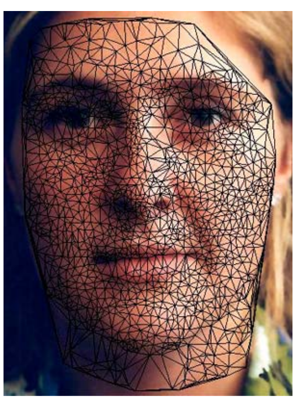


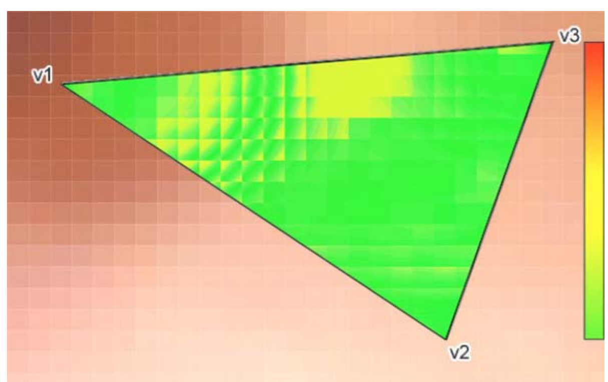
Geometric accuracy does not imply colour accuracy!

Laboratory of Motion Analysis & Virtual Reality, MAVR

http://homes.dsi.unimi.it/~borghese

Recursive re-tiling



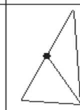
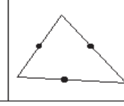
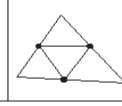


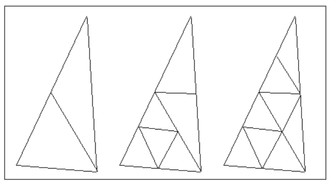
A limited chromatic error can be guaranteed.

Laboratory of Motion Analysis & Virtual Reality, MAVR

http://homes.dsi.unimi.it/~borghese

Metodologia di suddivisione del triangolo

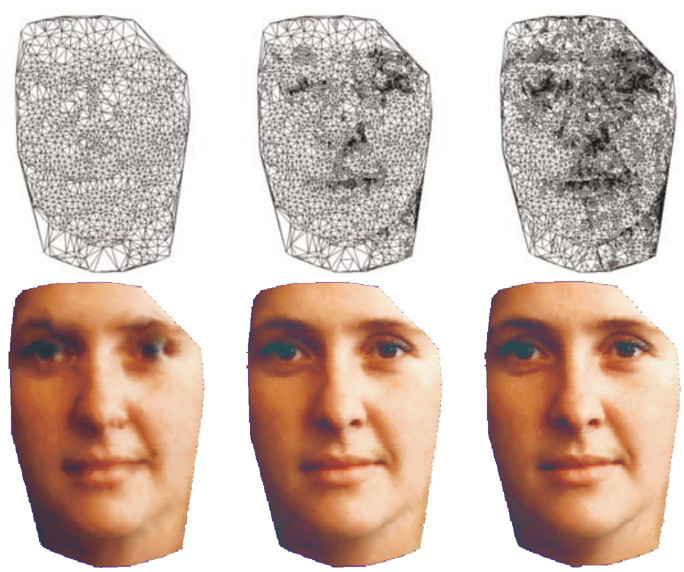
Classe	Condizione sugli angoli	Condizione sui lati	Classificazione	Selezione dei lati	Esempio	Suddivisione
1	Due angoli > 70°	Due lati > 146% del terzo lato	Isoscele alto	Il lato più lungo è diviso nel suo punto medio		
2	Due angoli < 45°	Due lati < 71% del terzo lato	Isoscele basso	Il lato più lungo è diviso nel suo punto medio		
3	Altrimenti	Altrimenti	Quasi–equilatero	Ogni lato è diviso nel suo punto medio		



La suddivisione del triangolo porta ad una graduale regolarizzazione della forma.

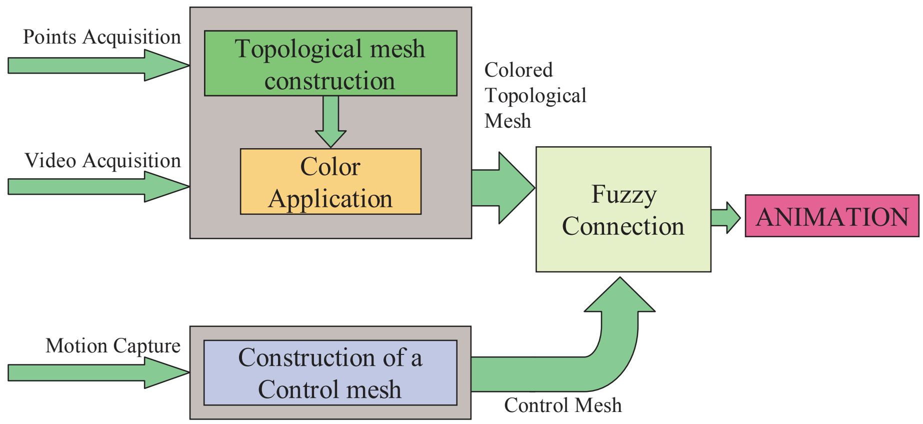
Re-tiling technique

Computational time < 1s



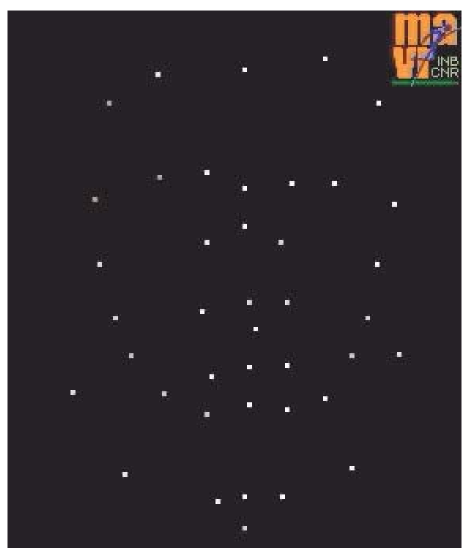
2,505 polygons 4,307 polygons 10,513 polygons

General schema



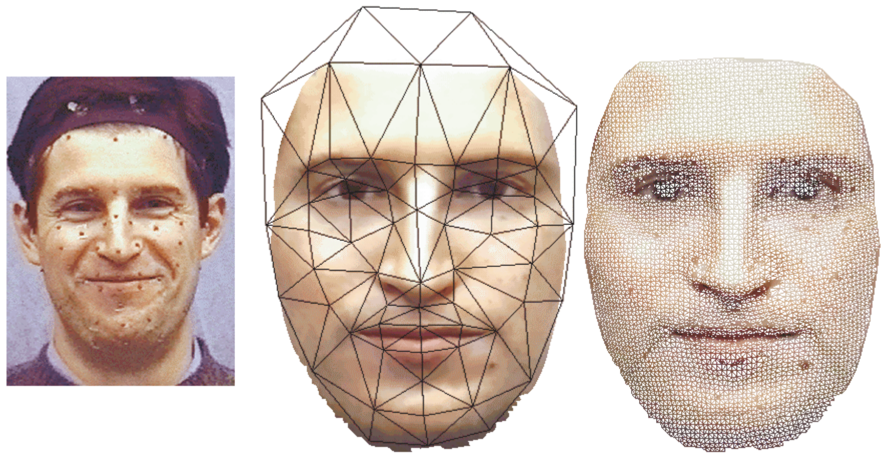
Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese

Two-layers Animation



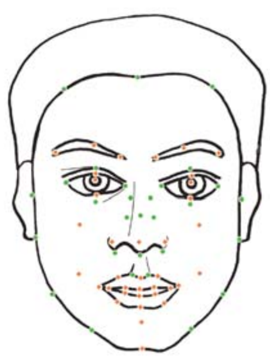
Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese

Two-layers technique



•Deformation of a topological mesh induced by a control mesh.
•The control mesh connects the marker points.

Markers disposition

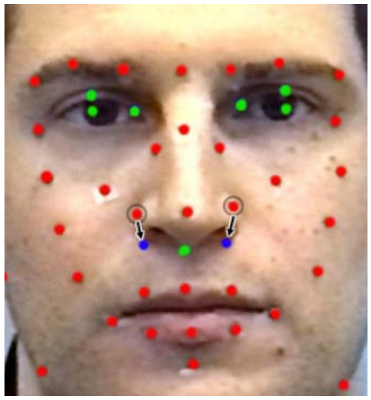


Position of the feature points
according to MPEG-4 standard:
◆ principali
● secondari

Problems with:
Eyes and tongue.
Nose basis (visibility).

Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese

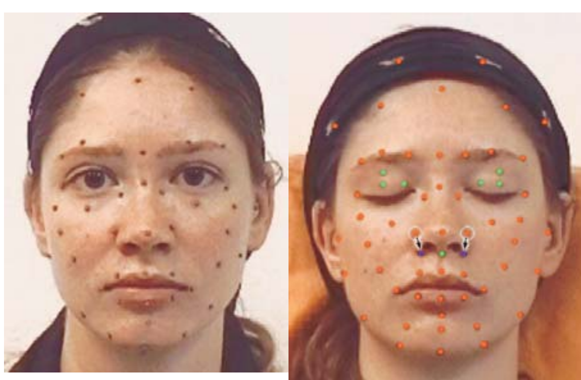
Real and Virtual markers

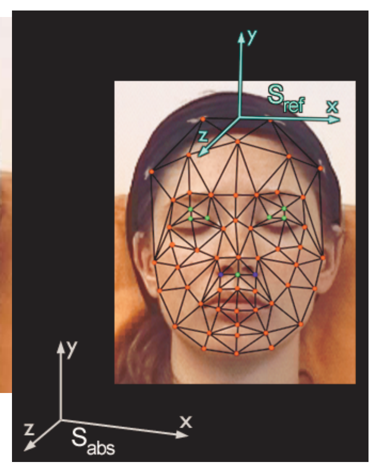


- Real markers (51)
- Virtual markers solid with the head (7)
- Virtual markers solid with other markers (2)

Total: 60 marker

Construction of the Control Mesh



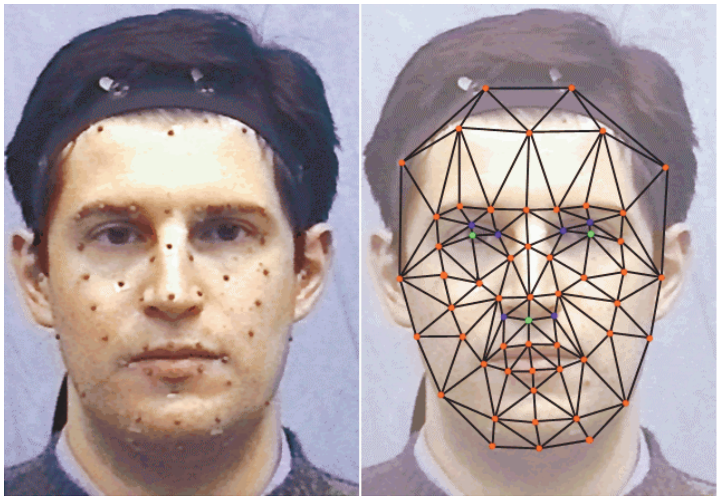


47 markers on the skin:
- *Problems with:*
  Eyes and tongue.
  Nose basis (visibility).

4 markers on an elastic band:
To identify a local Reference Frame (LRF).

- 51 Markers acquired (cf. MPEG-4 specifications).
- 7 virtual markers defined through the LRF (green).
- 2 Virtual markers defined through Real Markers (blue).
- 56 control points for the mesh + 4 for LRF.

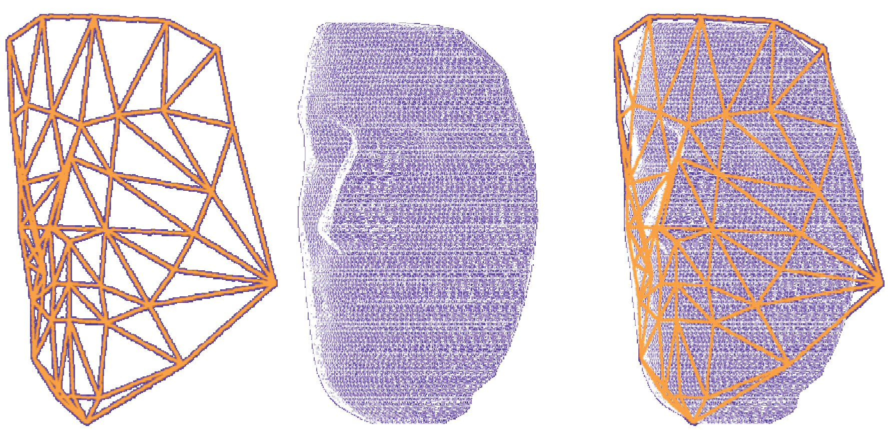
Construction of the Control Mesh



Laboratory of Motion Analysis & Virtual Reality, MAVR | http://homes.dsi.unimi.it/~borghese

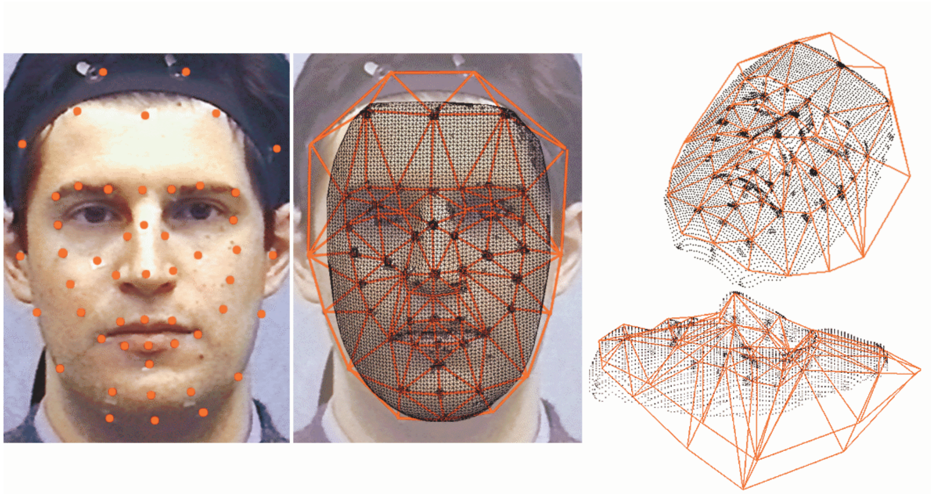
The topological and control mesh

Control mesh | Topological mesh | The 2 meshes aligned



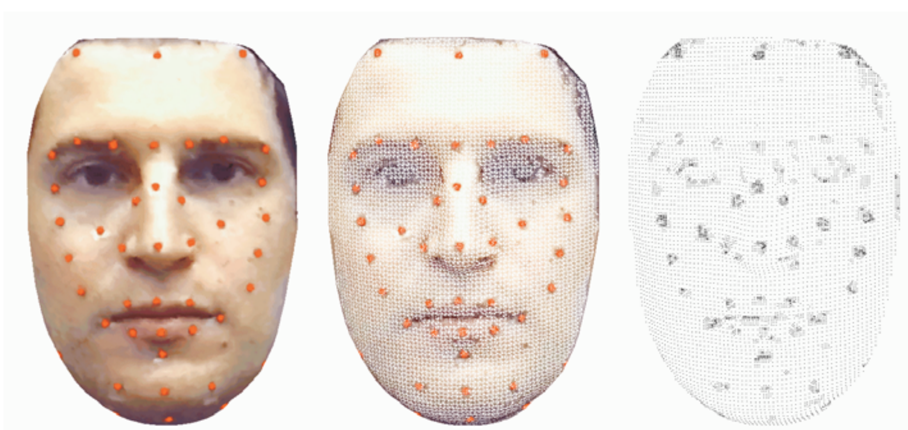
Laboratory of Motion Analysis & Virtual Reality, MAVR | http://homes.dsi.unimi.it/~borghese

Aligning the control and topological meshes

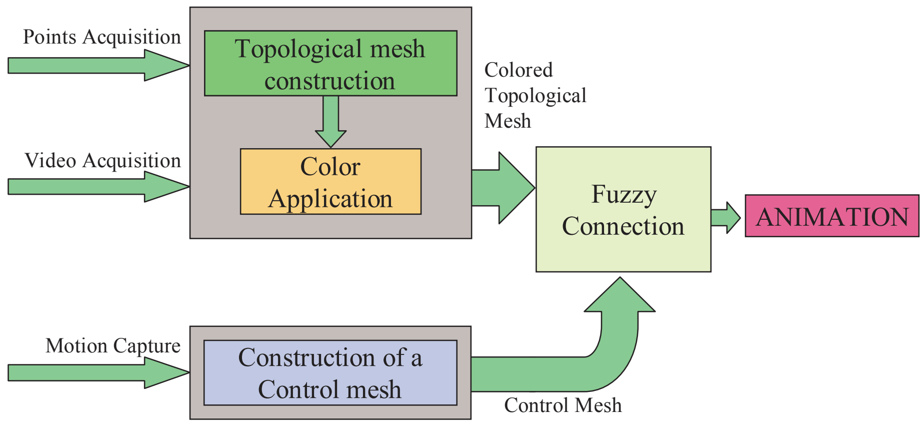


•The vertexes of the control mesh are superimposed to those of the topological mesh.
•The vertexes on the coloured topological mesh can be made more evident by using the adaptive tessellation procedure.

Markering the markers on the topological mesh



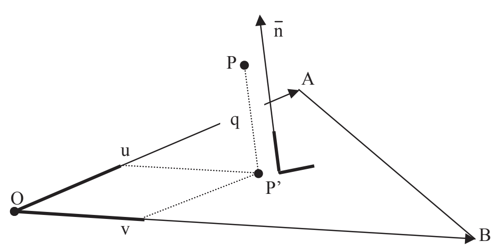
Towards animation



Laboratory of Motion Analysis & Virtual Reality, MAVR

http://homes.dsi.unimi.it/~borghese

Connection through intrinsic coordinates



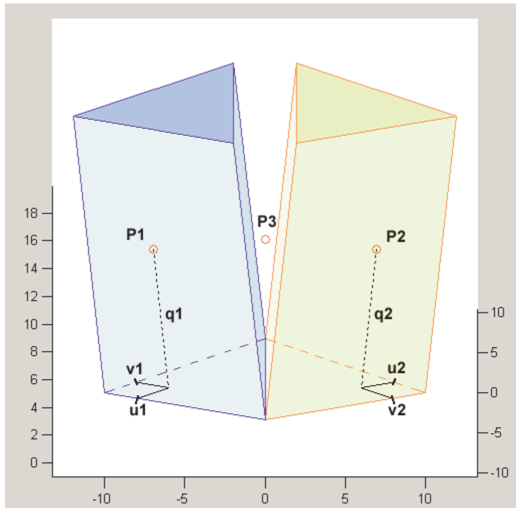
$$\mathbf{P} = \mathbf{O} + \overline{\mathbf{OB}} \bullet v + \overline{\mathbf{OA}} \bullet u + \vec{\mathbf{n}} \bullet q$$

u, v and q are the intrinsic coordinates of P

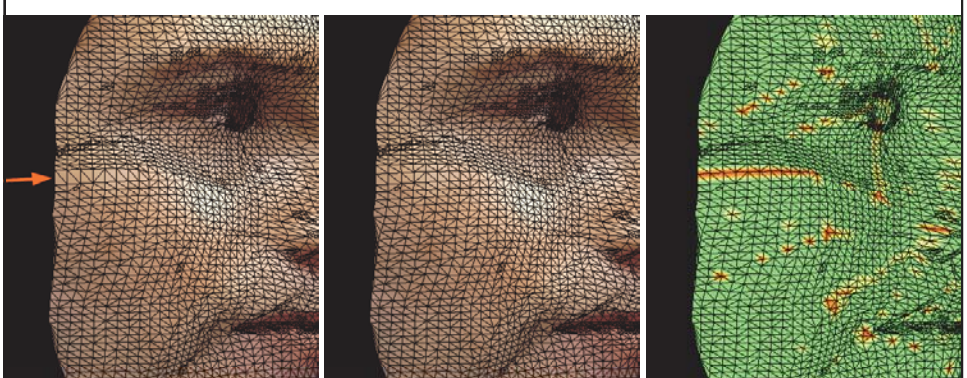
Laboratory of Motion Analysis & Virtual Reality, MAVR

http://homes.dsi.unimi.it/~borghese

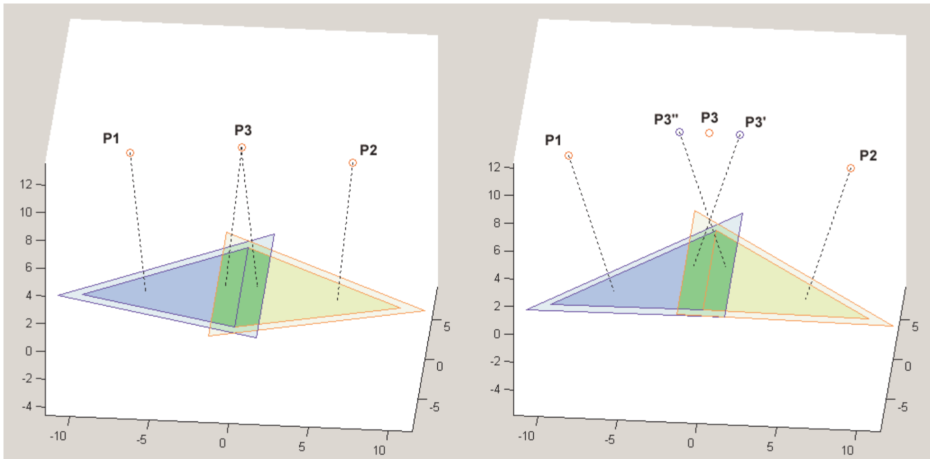
Determination of the proper control triangle



Stretches can form



Fuzzy association at the borders

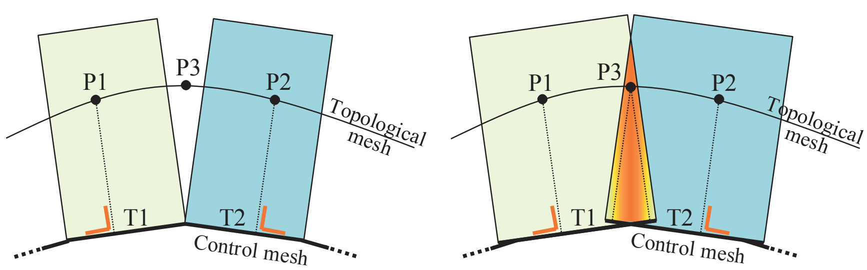


Connection between topological and control meshes

$\forall P_i$ of the topological mesh:

1) Determination of the tringle on which it is projected.
2) Computation of the intrinsic coordinates (for T_1 and T_2).

Problem at the border of the control tringles due to linear approximation:



$P1 \perp T1 \in T1 \quad P3 \perp T1 \notin T1$
$P2 \perp T2 \in T2 \quad P3 \perp T2 \notin T2$

No projection for point P3.

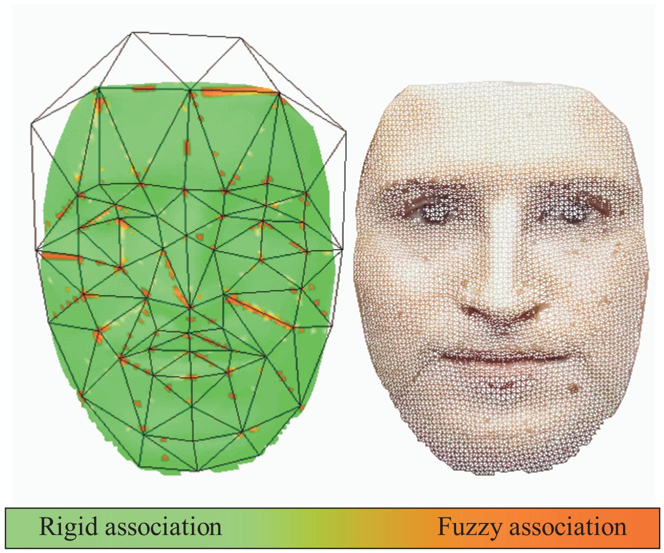
$P1 \perp T1 \in T1 \quad P3 \perp T1 \in T1$
$P2 \perp T2 \in T2 \quad P3 \perp T2 \in T2$

Two projections for point P3.

The solution is to give a fuzzy assignment at the borders.

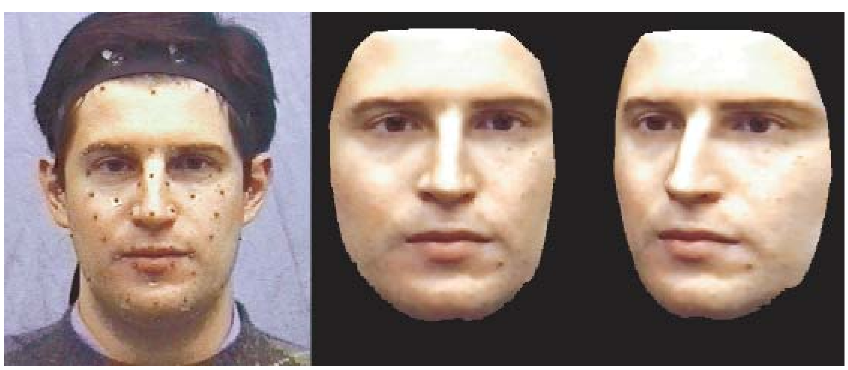
Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese

Result on fuzzy connection



Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese

Results: anger



Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese

Results: surprise

Results: disgust

Results: happiness

Real-time Animation

•Compact and realistic mesh, adequate for real-time animation.

•Future developments: Insert biomechanics “rules” into the exterior reproduction.

• Attract “good” psychologists to work with.

• Develop better compact models for virtual interactions.