Algorithms for Nonrigid Reconstruction and Recognition

Francesc Moreno-Noguer

Institut de Robòtica i
Informàtica Industrial

Institut de Robòtica i Informàtica Industrial

28 Doctors

10 Technicians

37 Doctorands

14 Support personnel

>10 Robots

Perception and Manipulation Group

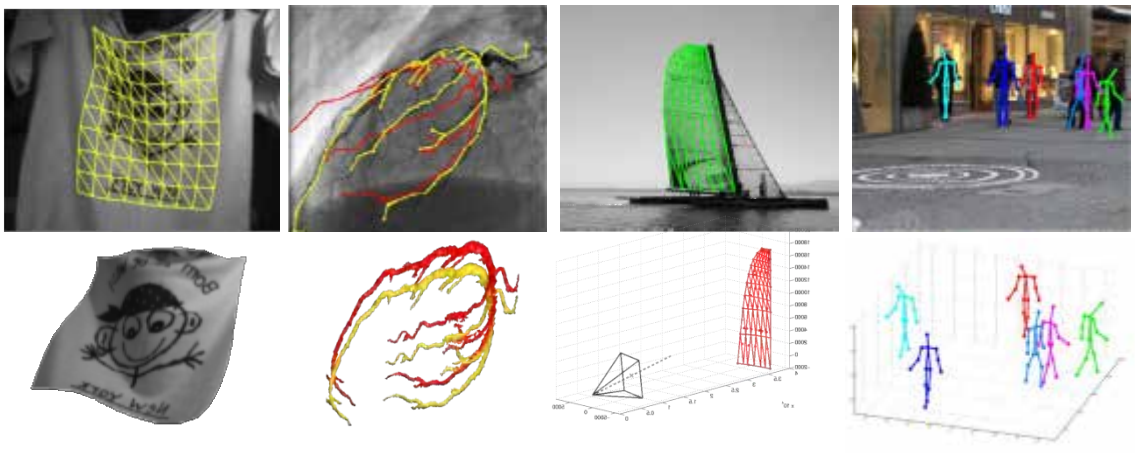
Robot manipulators in human environments (service, assistive, domestic, industrial)

Perception and Manipulation Group

- Easy to program by non-experts
- Safe for people
- Tolerant to noisy perceptions and inaccurate actions
- Adaptable
- Able to perceive and manipulate deformable objects

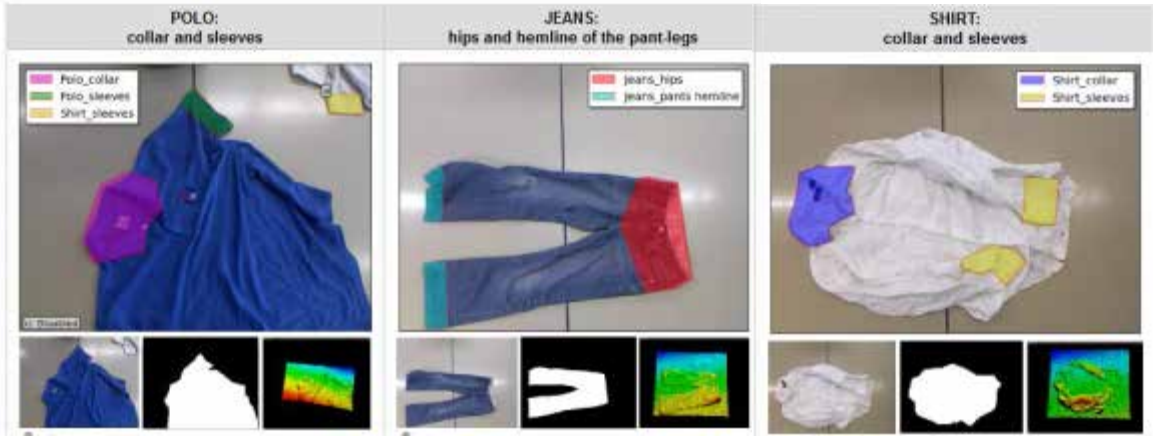
Perception

Non Rigid Shape Estimation

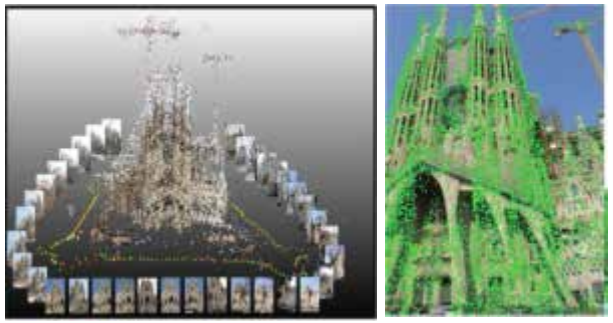


Enhanced Descriptors

Non-Rigid Shape Recognition

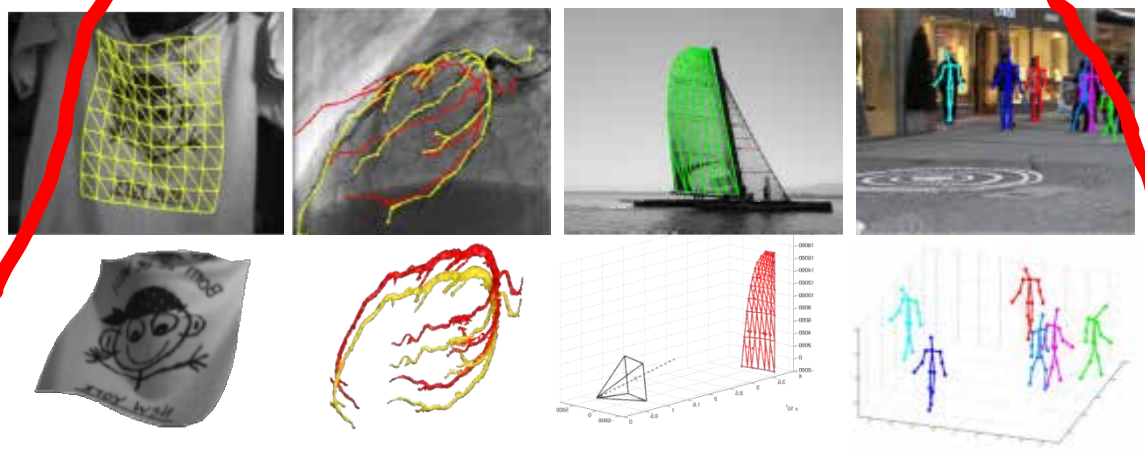


Pose Estimation



Perception

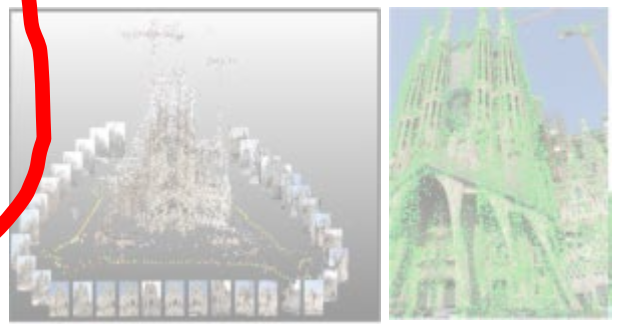
Non Rigid Shape Estimation



Non-Rigid Shape Recognition

Enhanced Descriptors

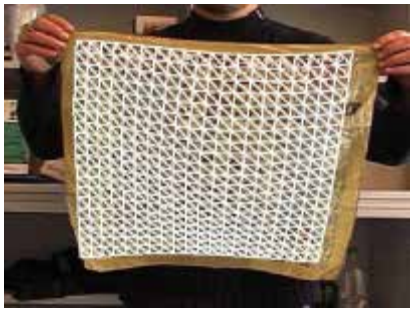
Pose Estimation

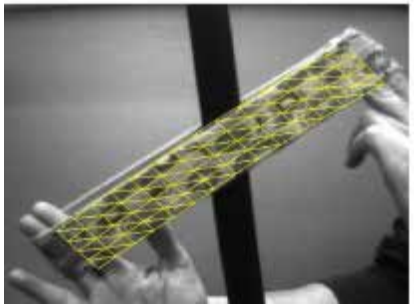


Outline

Non-Rigid Detection

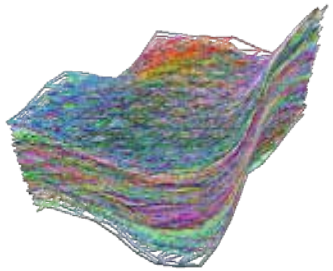
(ECCV'08, CVPR'09)

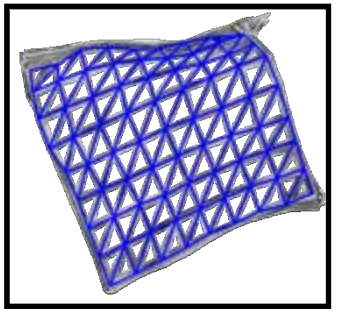




Limitations of Linear Formulations

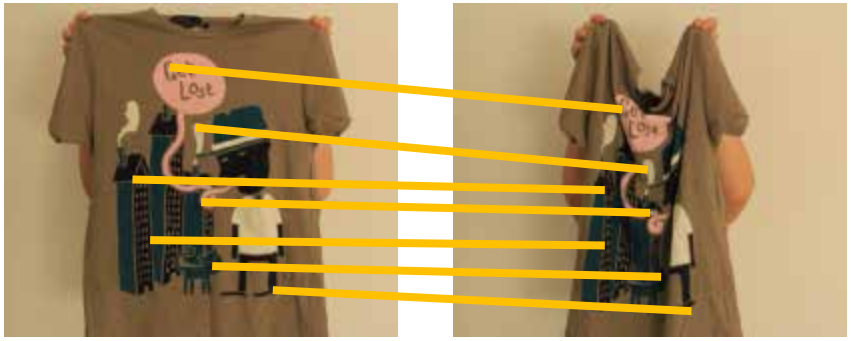
(ECCV'10, CVPR'12, PAMI'13)

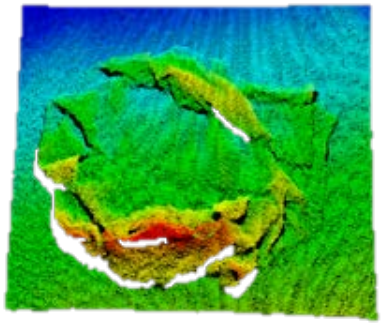


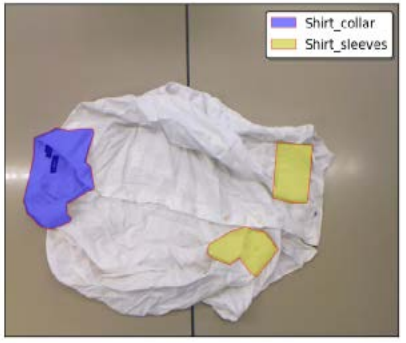


Non-Rigid Recognition

(CVPR'11, ICRA'12, IROS'13)



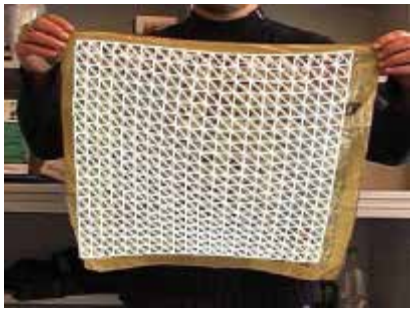


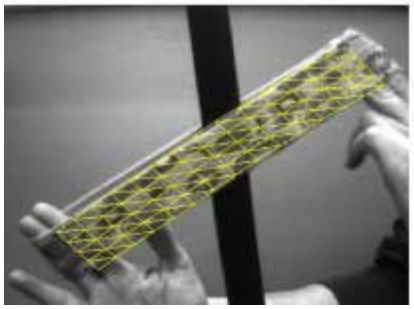


Outline

Non-Rigid Detection

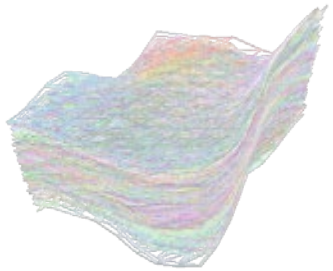
(ECCV'08, CVPR'09)

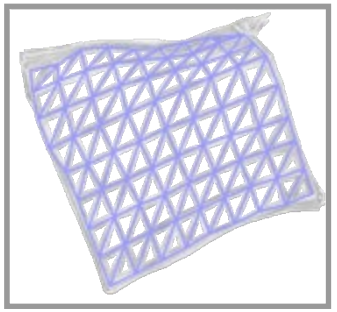




Limitations of Linear Formulations

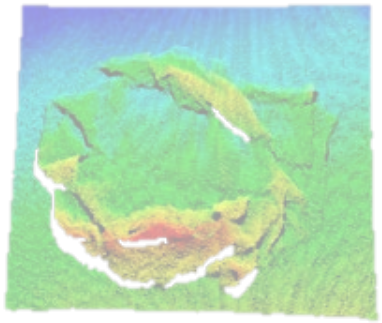
(ECCV'10, CVPR'12, PAMI'13)





Non-Rigid Recognition

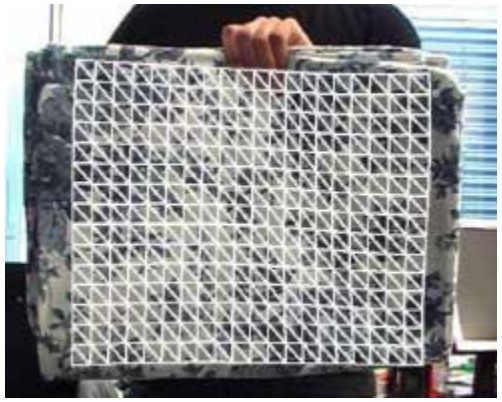
(CVPR'11, ICRA'12, IROS'13)



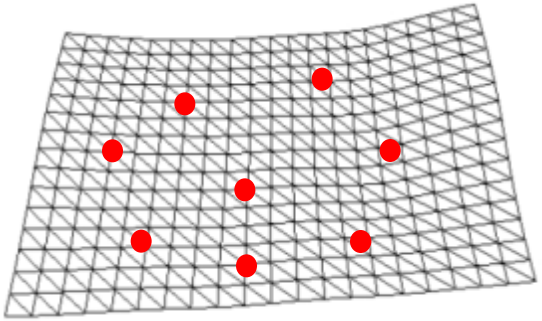
Non-Rigid Detection Problem

- **Given:** 2D/3D correspondences between **reference configuration** and **input image** + Camera internal parameters
- **We want:** Retrieve the 3D shape in the input image

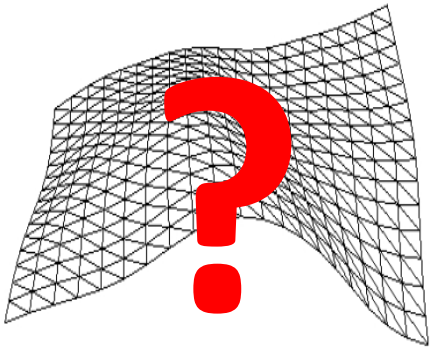
Reference Configuration



Input Image



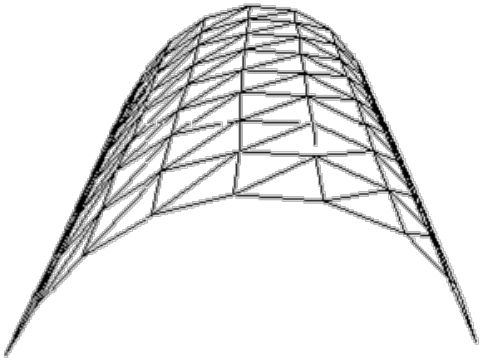
3D Points



In closed form!!

Linear Formulation

- The surface is a triangulated mesh with
  n_v vertices $\mathbf{v}_i = [x_i, y_i, z_i]^T$



- We want to recover $\mathbf{x} = [\, \mathbf{v}_1^T \ \ldots \ \mathbf{v}_{n_v}^T \,]^T$

- A correspondence is defined by
  - a 3D point given in barycentric coordinates (a_1, a_2, a_3)
  - a 2D location in the input image (u, v)

  } **Known**

- The 3D-to-2D projection can be written as

$$\mathbf{A}\,(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3) = k \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

Known calibration matrix ↑

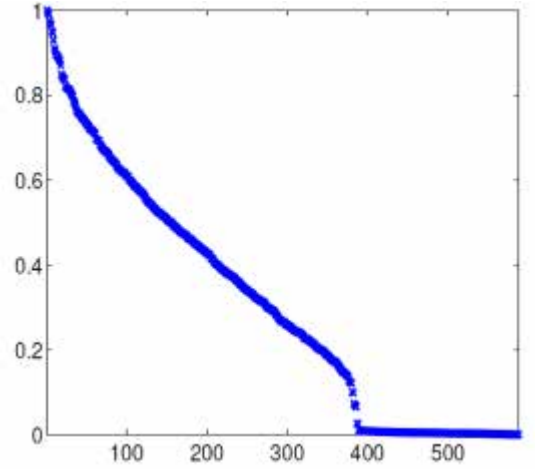
Projective scale parameter ↑

Linear Formulation

- n_C such correspondences yield the linear system

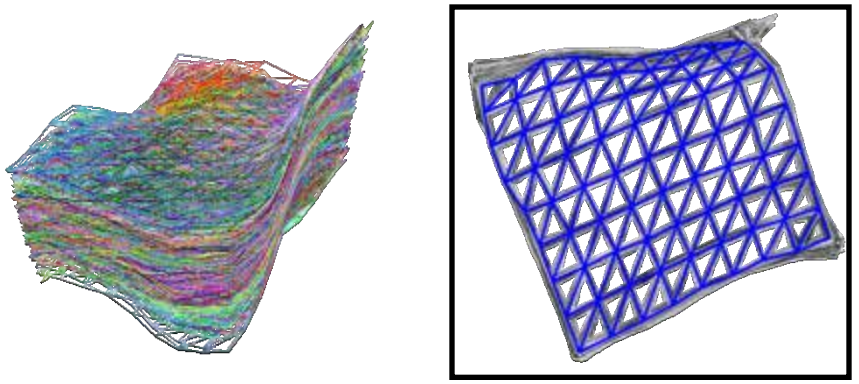
Constant matrix made of known coefficients → $\mathbf{Mx} = \mathbf{0}$

- Problem: In practice this is under-constrained



Eigenvalues of **M**

View ambiguity



Different shapes → same reprojection

We need additional constraints to disambiguate

Linear Deformation Model

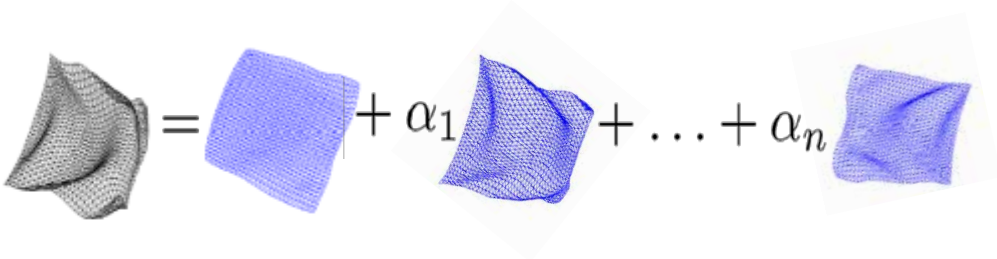
- Shape = linear combination of modes

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^{n_m} \alpha_i \mathbf{q}_i$$

Mean shape

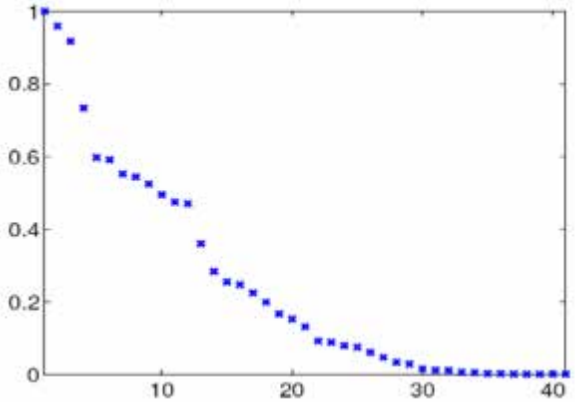
Unknown weights

Modes



- The correspondence problem becomes

$$\tilde{\mathbf{M}} \begin{bmatrix} \boldsymbol{\alpha} \\ 1 \end{bmatrix} = 0$$



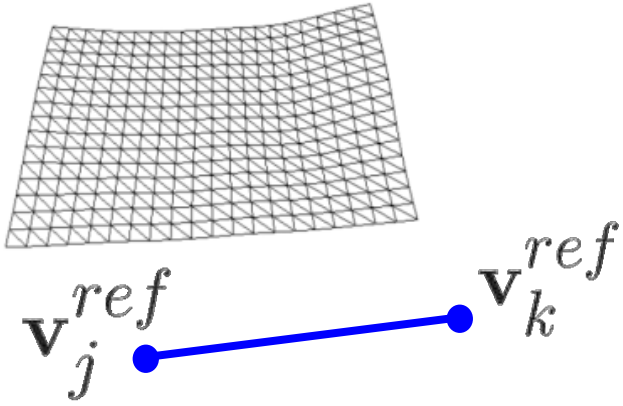
Eigenvalues of $\tilde{\mathbf{M}}$

Still under-constrained, but much less…

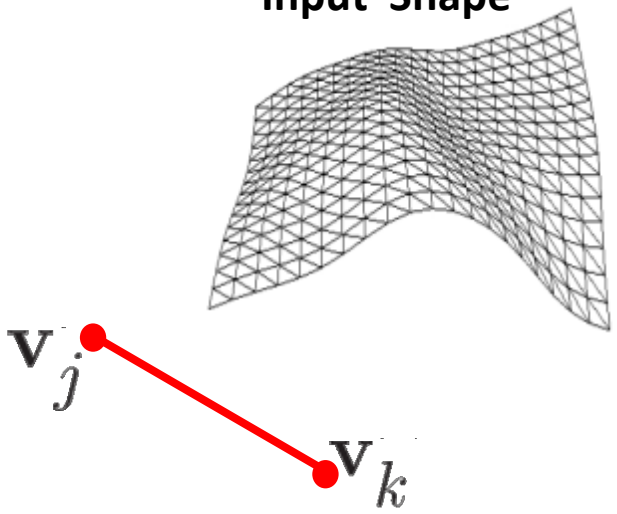
Inextensibility Constraints

- We add a distance constraint for each edge of the mesh

Reference Configuration

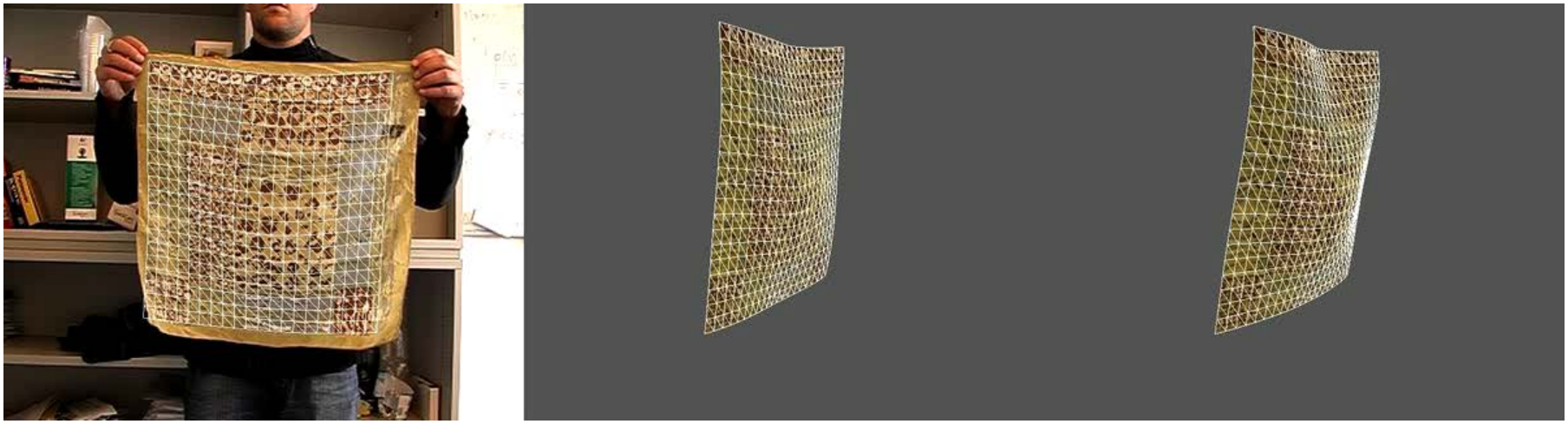


Input Shape



$$\underbrace{\|\mathbf{v}_j^{ref} - \mathbf{v}_k^{ref}\|^2}_{\text{Known distances}} = \underbrace{\|\mathbf{v}_j - \mathbf{v}_k\|^2}_{\text{Linear and quadratic constraints on the } \alpha\text{´s}}$$

Results

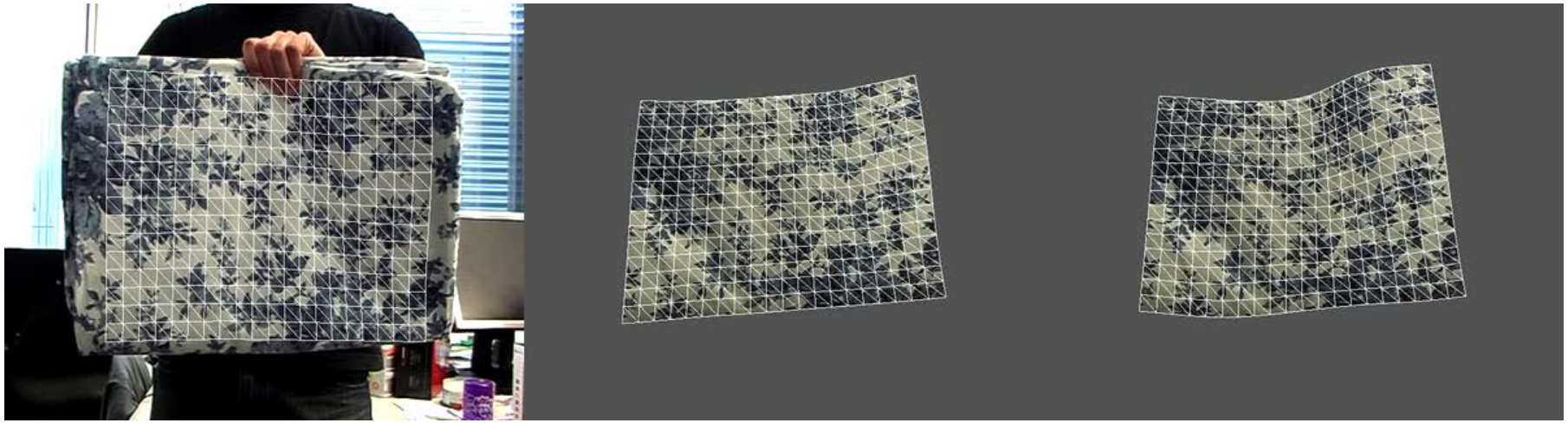


Overlaid mesh **Closed form solution** **After 5 iterations**

- Independent detection in every single frame
- Shaking due to some remaining ambiguities

[ECCV'2008]

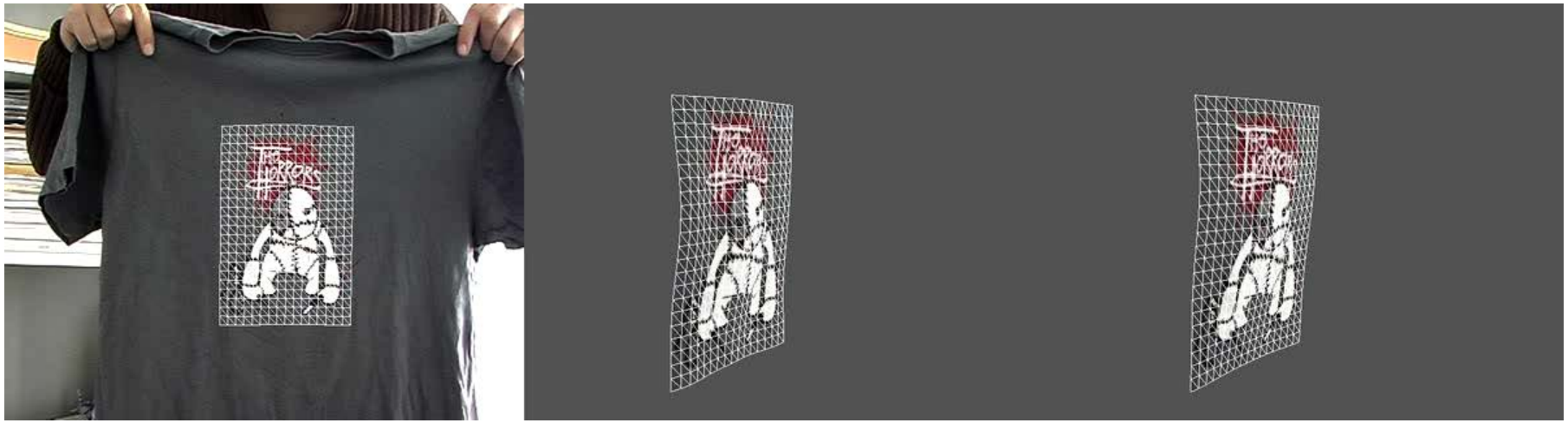
Results



Overlaid mesh | Closed form solution | After 5 iterations

[ECCV'2008]

Results



Overlaid mesh **Closed form solution** **After 5 iterations**

[ECCV'2008]

Results

Overlaid mesh **Closed form solution** **After 5 iterations**

- The use of a global model (PCA) gives robustness to occlusions.

[ECCV'2008]

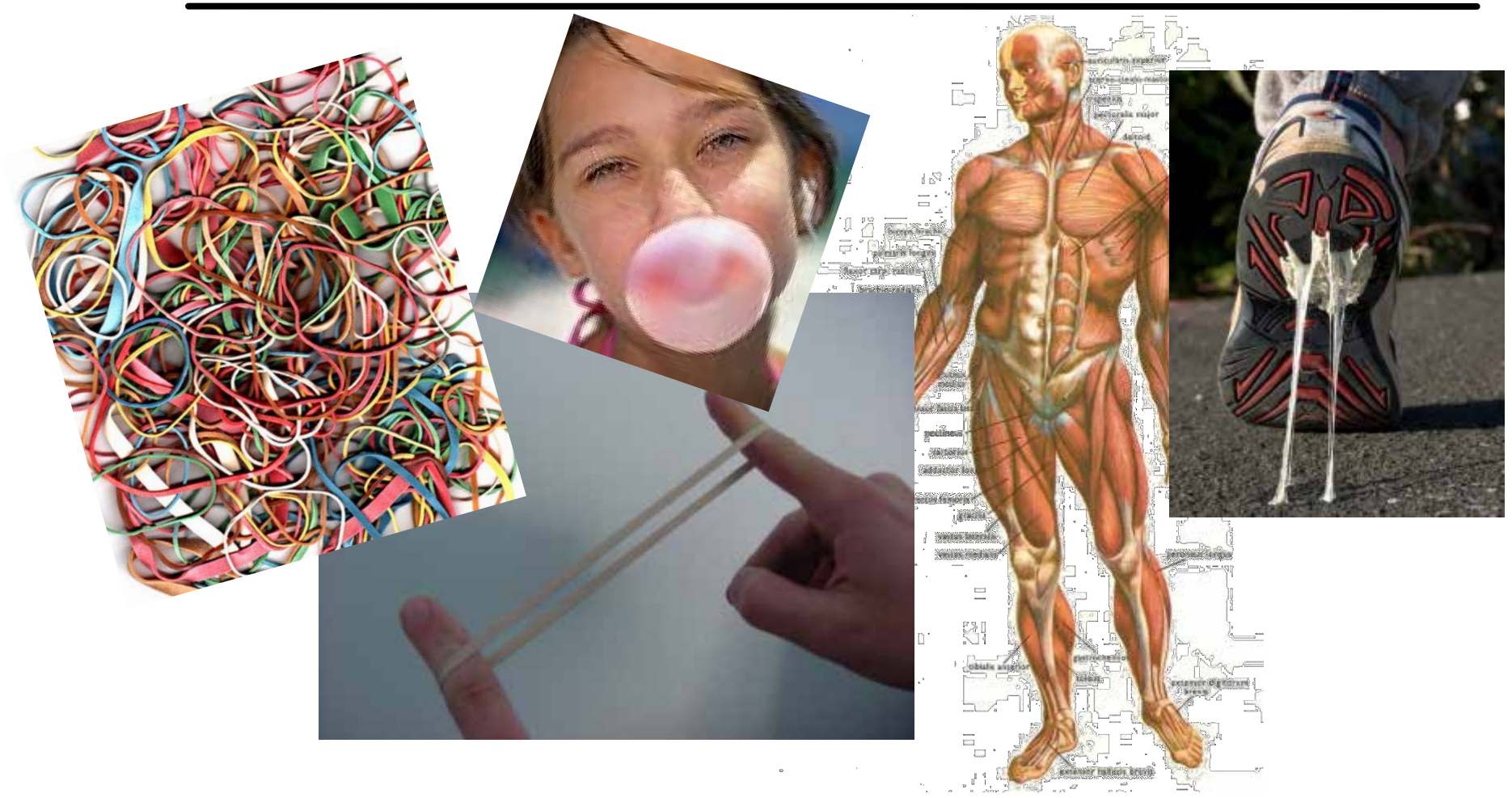
Conclusions & Limitations

- Closed-form solution to non-rigid 3D surface registration
  - + First closed form solution to that problem
  - + Does not require from initialization

- Limitations
  - – Still some ambiguities in the solution
  - – Inextensibility constraints → cannot handle stretchable surfaces

Tried to address this limitation

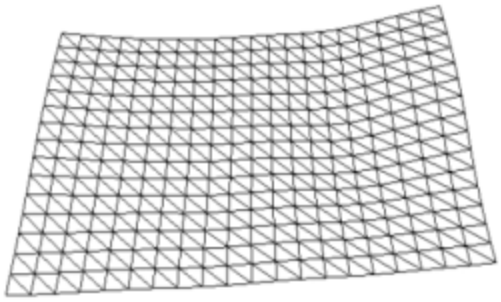
Detecting Elastic Surfaces



Many objects do not satisfy local inextensibility constraints

Removing Inextensibility Constraints

Reference Configuration

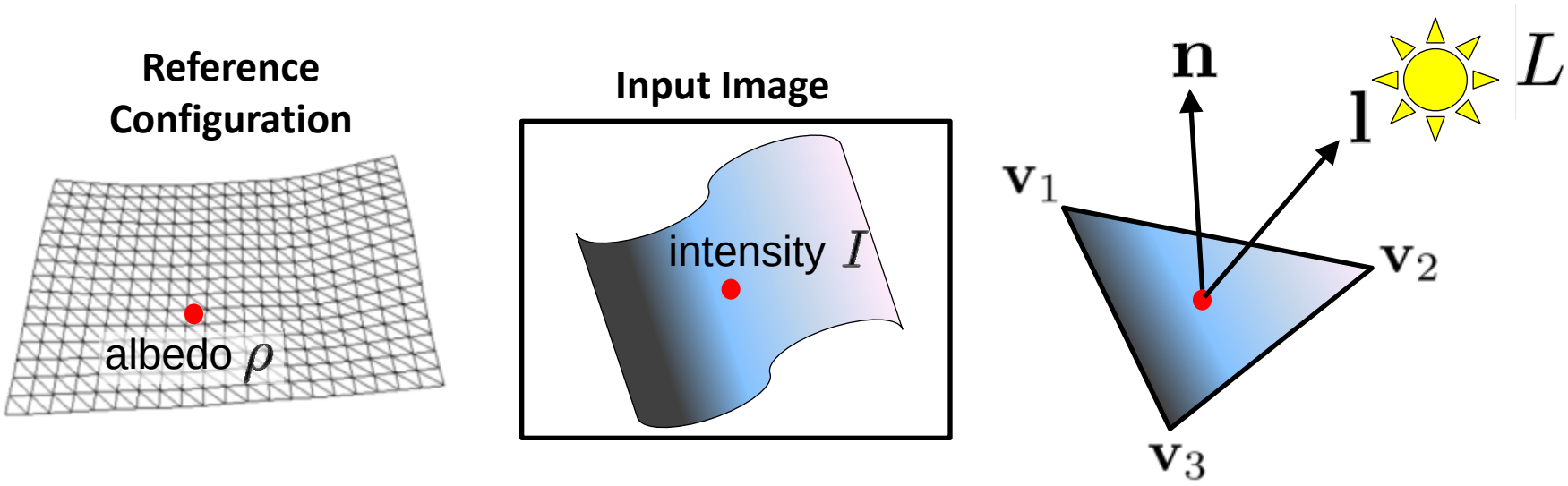


$$\|\mathbf{v}_j - \mathbf{v}_k\|^2 = \|\mathbf{v}_j^{ref} - \mathbf{v}_k^{ref}\|^2$$

Linear and quadratic terms on the α's

Known distances

Shading Constraints



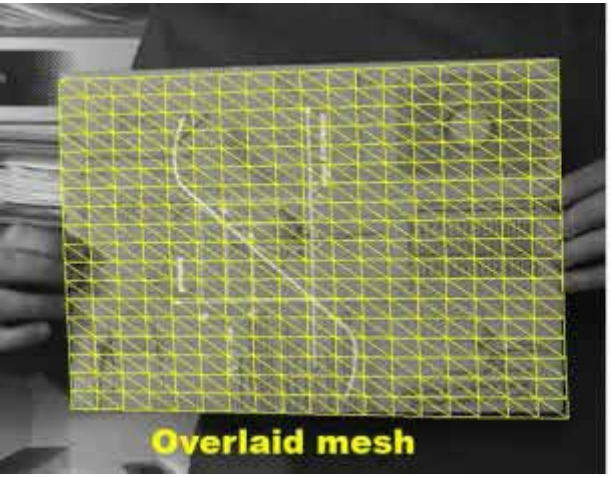
- Shading constraint (Lambertian model) : $I = \rho L(\mathbf{l} \cdot \mathbf{n})$
- $\mathbf{n}$ can be written in terms on vertices coordinates

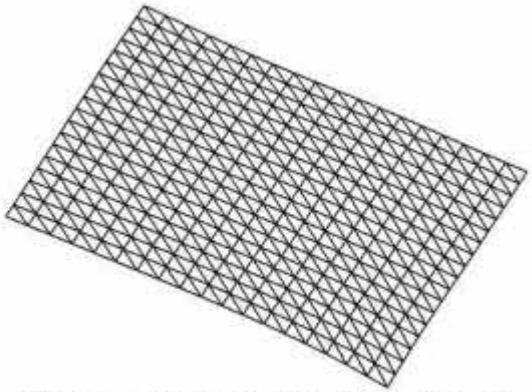
$$n_x = y_2 z_3 - y_2 z_1 - y_1 z_3 - z_2 y_3 + z_2 y_1 + z_1 y_3$$

$$n_y = z_2 x_3 - z_2 x_1 - z_1 x_3 - x_2 z_3 + x_2 z_1 + x_1 z_3$$

$$n_z = x_2 y_3 - x_2 y_1 - x_1 y_3 - y_2 x_3 + y_2 x_1 + y_1 x_3$$

Results: Inelastic Deformation





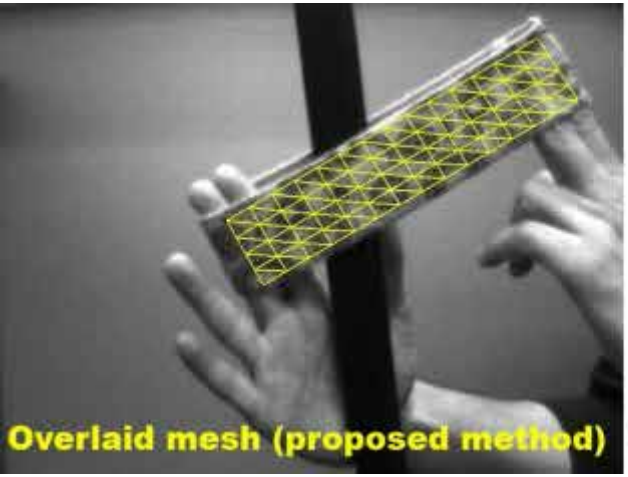
3D mesh + Light directions

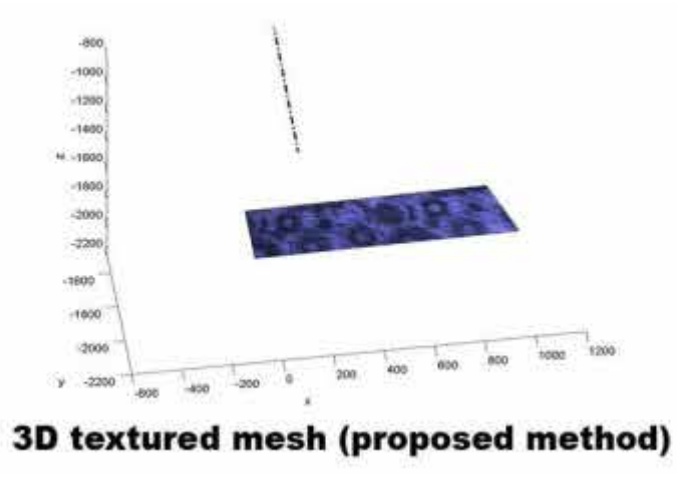
Synthesized textured view

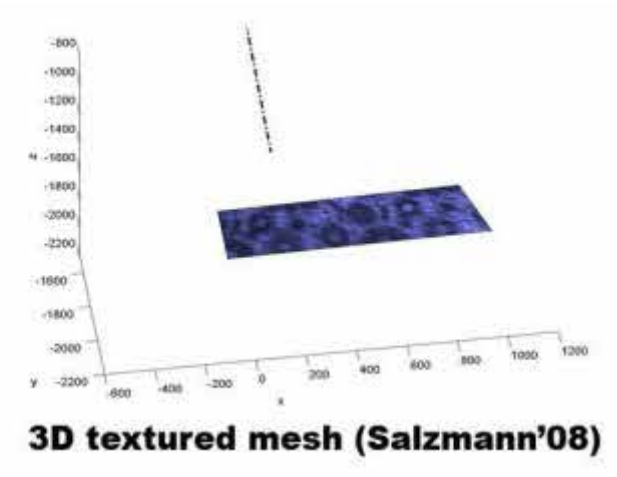
- Independent detection in every single frame
- Shape + Lighting parameters

[CVPR'2009]

Results: Elastic Deformation





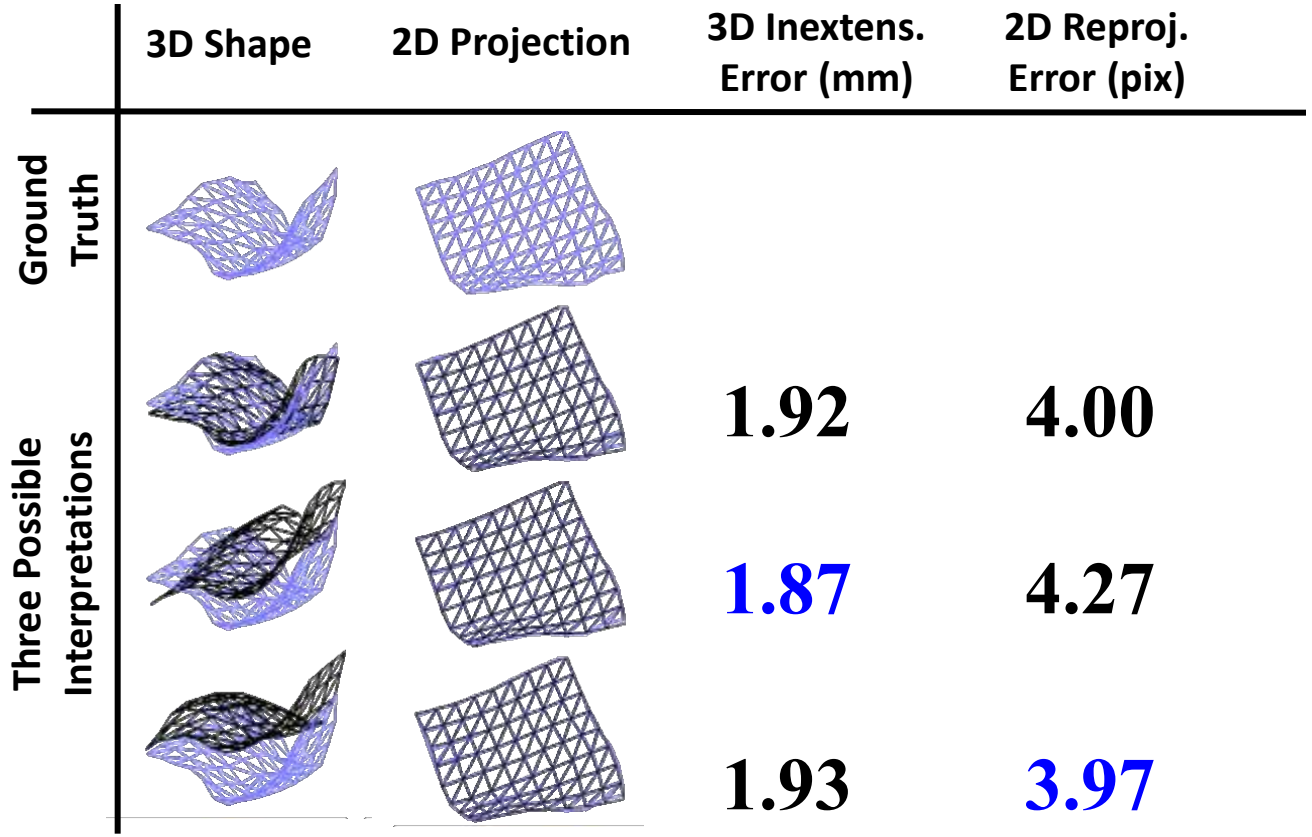


- Inextensibility constraints 'believe' the mesh is approaching the camera.

[CVPR'2009]

Limitations of Linear Formulations

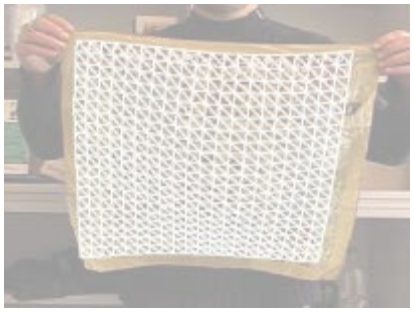
- Limitation: Still some ambiguities in the solution.
  - *Reprojection* and *inextensibility* constraints are not sufficient to disambiguate.

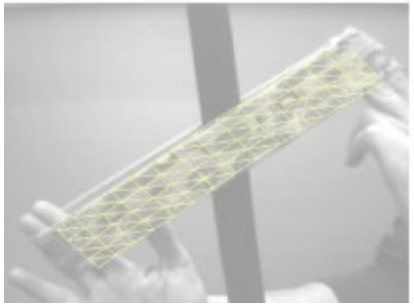


Outline

Non-Rigid Detection

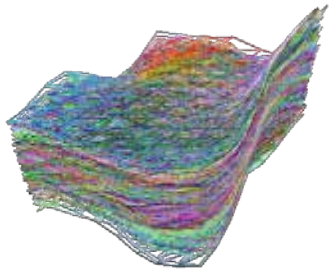
(ECCV'08, CVPR'09)

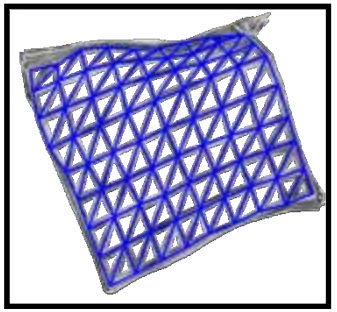




Limitations of Linear Formulations

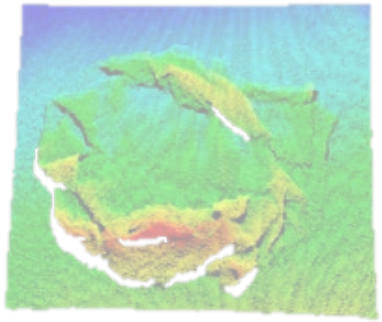
(ECCV'10, CVPR'12, PAMI'13)





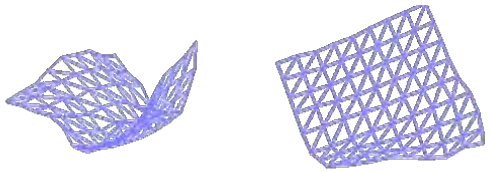
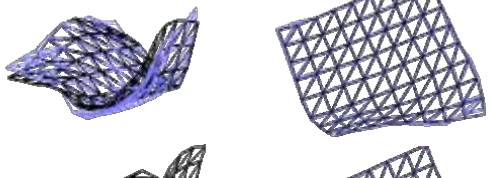
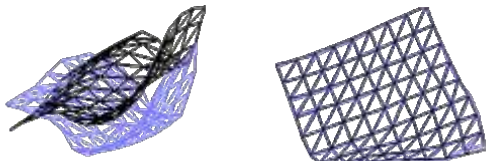
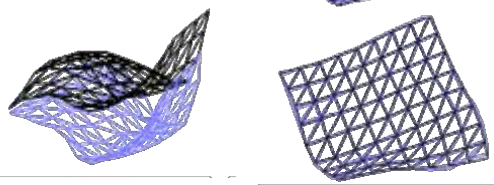
Non-Rigid Recognition

(CVPR'11, ICRA'12, IROS'13)



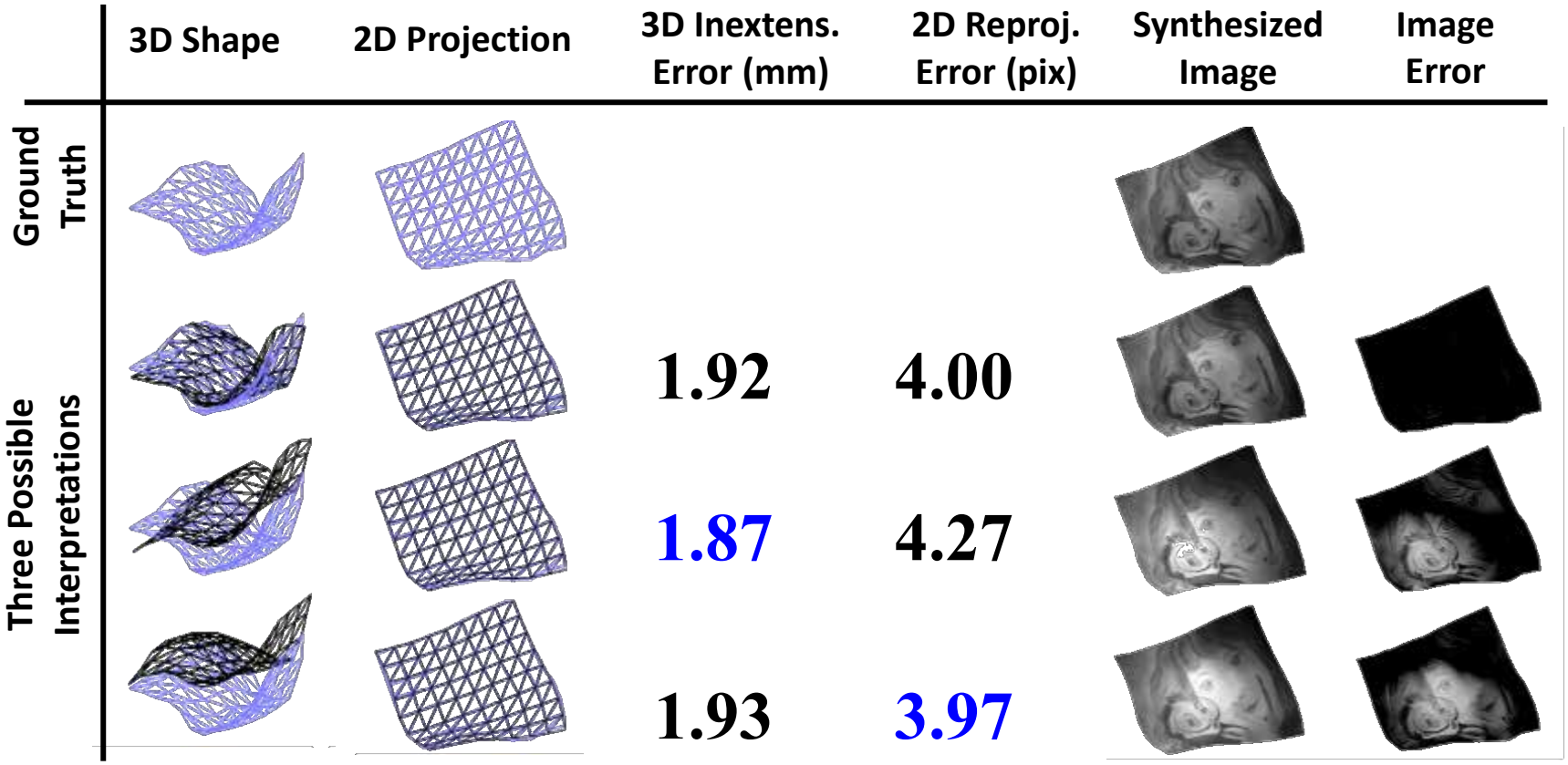
Exploring Ambiguities

- We will explore the set of ambiguous solutions.

	3D Shape	2D Projection	3D Inextens. Error (mm)	2D Reproj. Error (pix)
Ground Truth				
Three Possible Interpretations			1.92	4.00
			1.87	4.27
			1.93	3.97

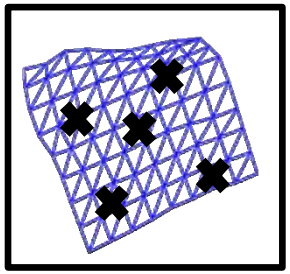
Exploring Ambiguities

- We will explore the set of ambiguous solutions.
- We will then apply more discriminative constraints to retain one single solution.



Proposing Candidate Shapes

- **Goal:** Retrieving a set of shapes that are correctly reprojected and satisfy inextensibility constraints.
- We start from the linear formulation of the problem.



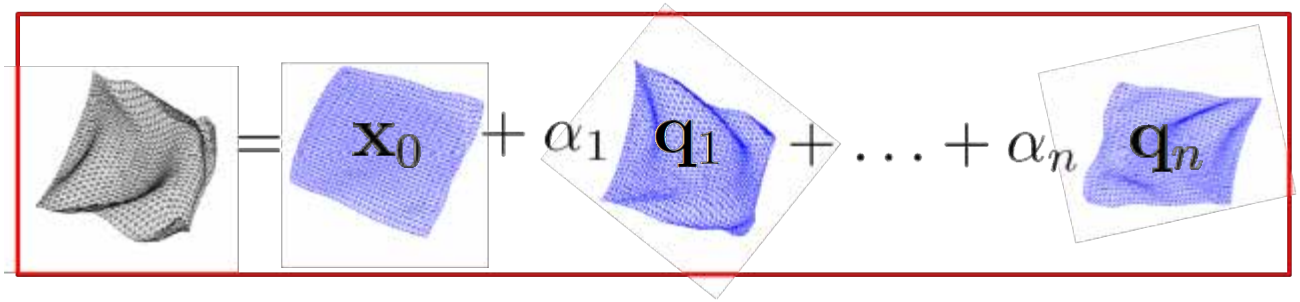
2D correspondences

Mean Shape

$$\mathbf{M}\mathbf{Q}\boldsymbol{\alpha} + \mathbf{M}\mathbf{x}_0 = 0$$

Deformation Modes

Modal Weights (Define the Shape)



- It defines a mapping between the 2D coordinates and the shape space.

Proposing Candidate Shapes

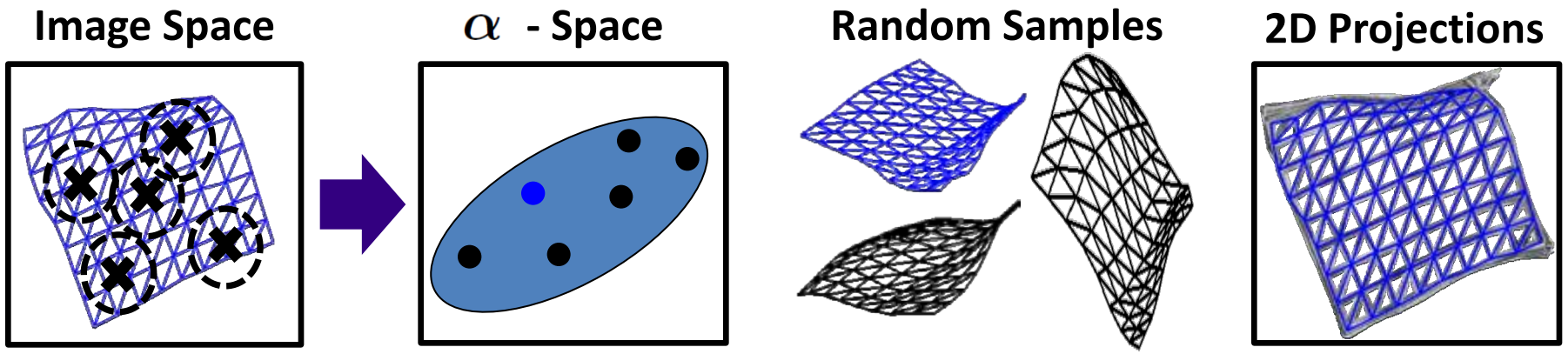
- Assume 2D locations $\mathbf{u}_i$ are normally distributed

$$\mathbf{u}_i \sim \mathcal{N}(\mathbf{u}_i, \Sigma_{\mathbf{u}})$$

- We propagate the error to the modal weights space

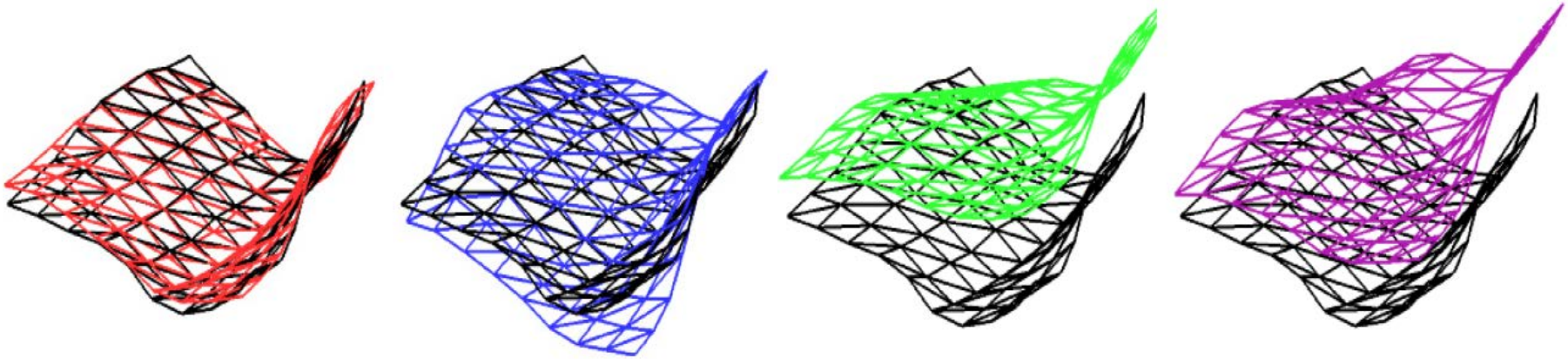
$$\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \Sigma_{\boldsymbol{\alpha}})$$

- Sample the $\boldsymbol{\alpha}$-space and retain the shapes that best satisfy reprojection and intextensibility constraints.

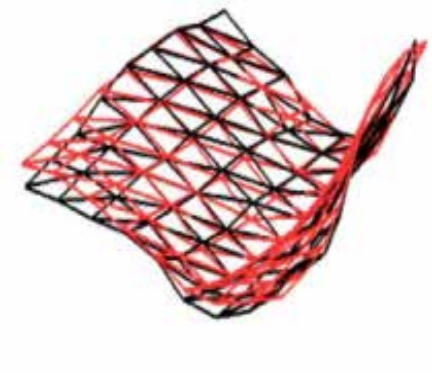


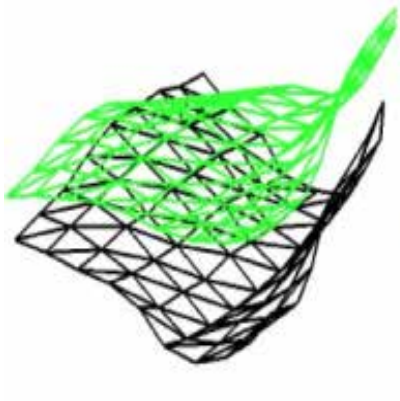
Using Shading to Disambiguate

- Set of ambiguous shapes

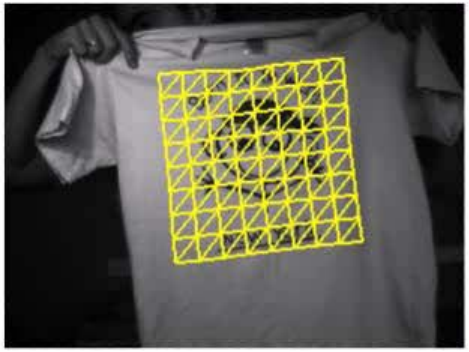


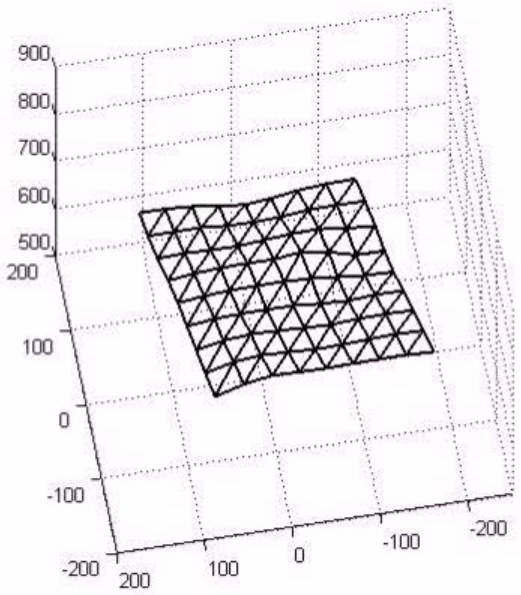
- Render the input image and compute shading error

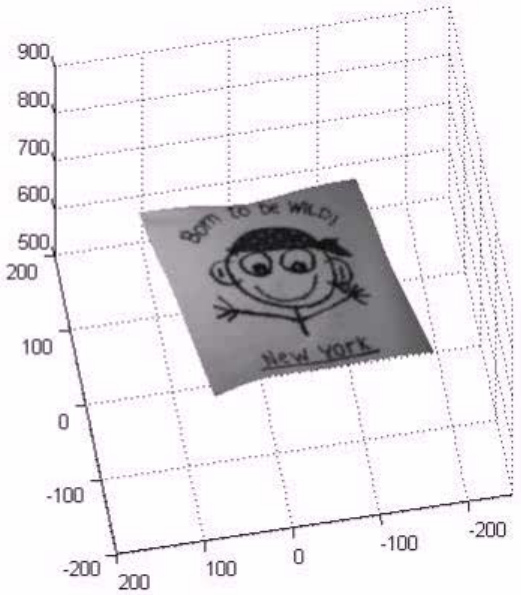




Results



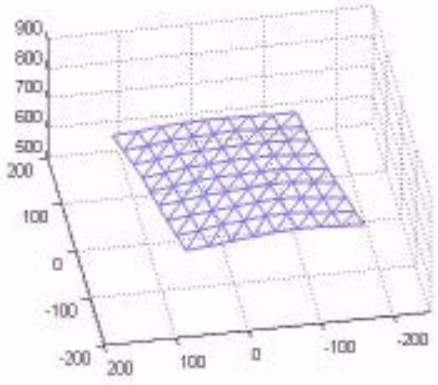




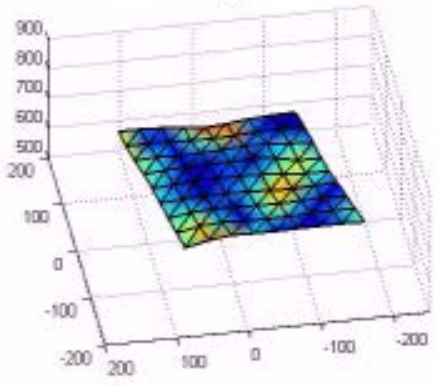
[ECCV 2010, PAMI 2013]

Results

Ground Truth

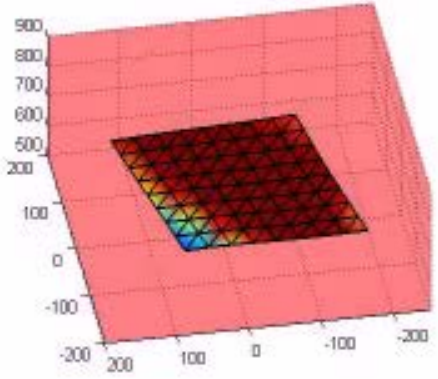


Our Approah

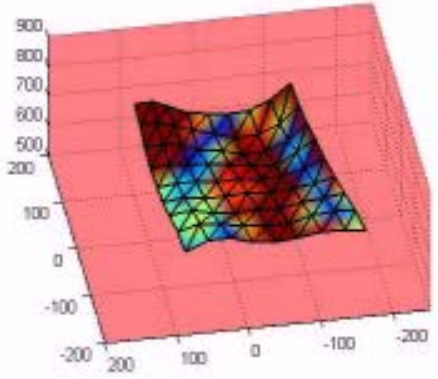


2D Projection
(Our Approach)

Salzm09



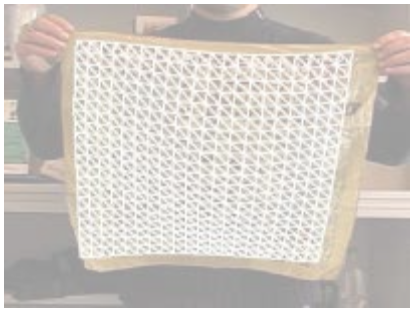
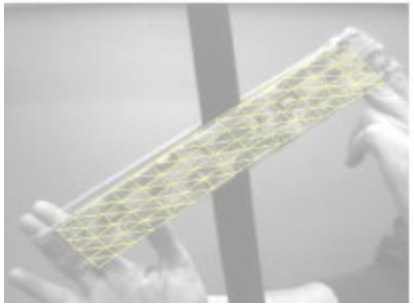
Moreno09



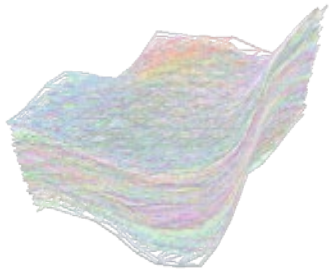
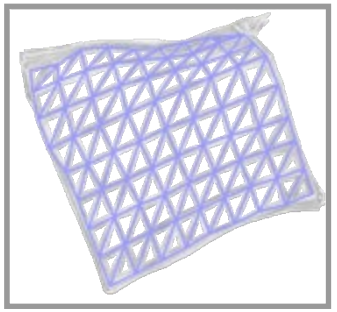
[ECCV 2010, PAMI 2013]

Outline

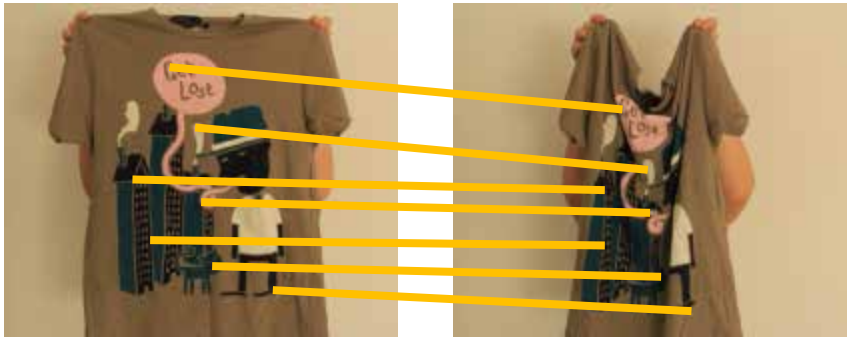
Non-Rigid Detection
(ECCV'08, CVPR'09)

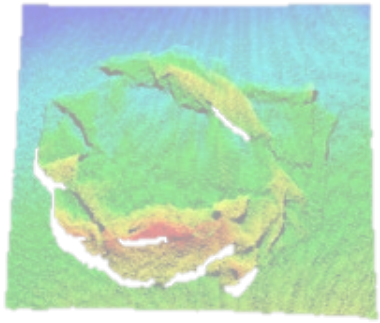



Limitations of Linear Formulations
(ECCV'10, CVPR'12, PAMI'13)

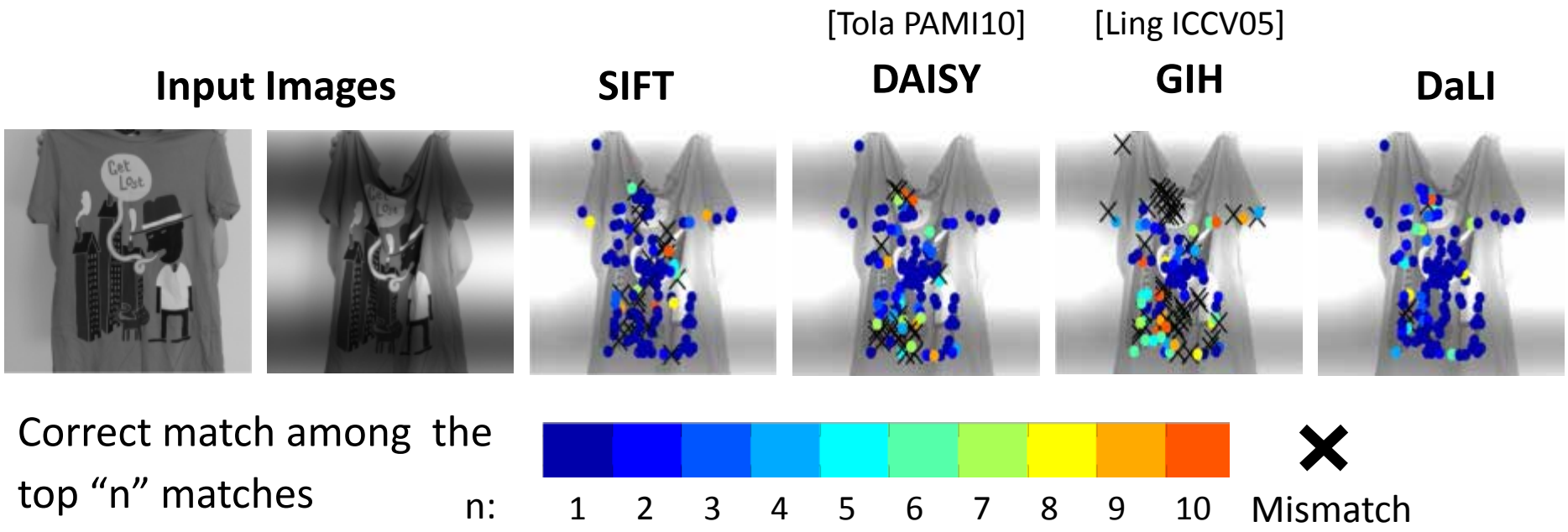
Non-Rigid Recognition
(CVPR'11, ICRA'12, IROS'13)






What about correspondences?

- Problem: Match points of interest under:
  - Non-rigid deformations
  - Photometric changes
- DaLI: Deformation and Light Invariant descriptor



Representing Patches as Surfaces

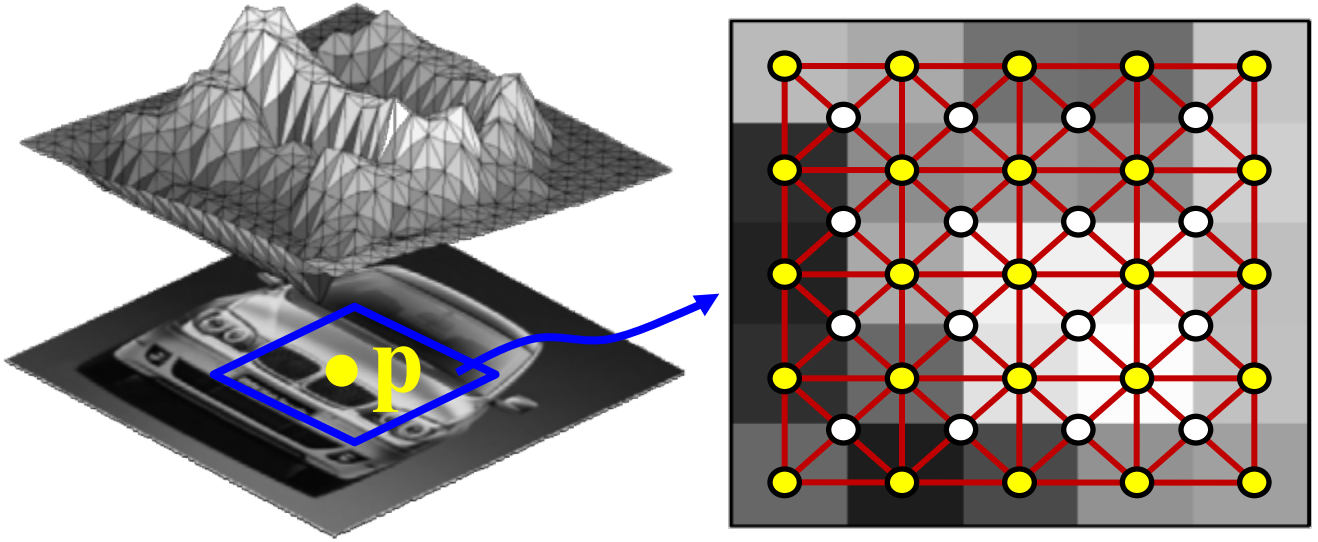
- DaLI: Uses heat diffusion theory to describe 2D patches.
- Heat diffusion geometry has been used for non-rigid 3D shape recognition.

Bronstein ECCV10

Heat diffusion is invariant to non-rigid deformations

DaLI Computation

1. Represent patches as triangulated surfaces



$$\mathbf{x} \longrightarrow (x, y, \beta I(x,y))$$

2. Compute the Laplace-Beltrami operator:

$$\mathbf{L} = \mathbf{A}^{-1}\,\mathbf{M}$$

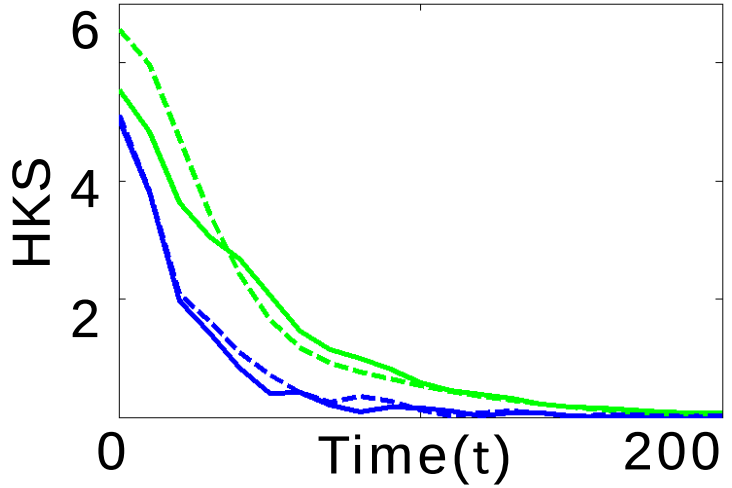
Positive semidefinite matrix with the structure of the mesh

Diagonal matrix with $\mathbf{A}_{ii}$ ~ area i-th vertex

DaLI Computation

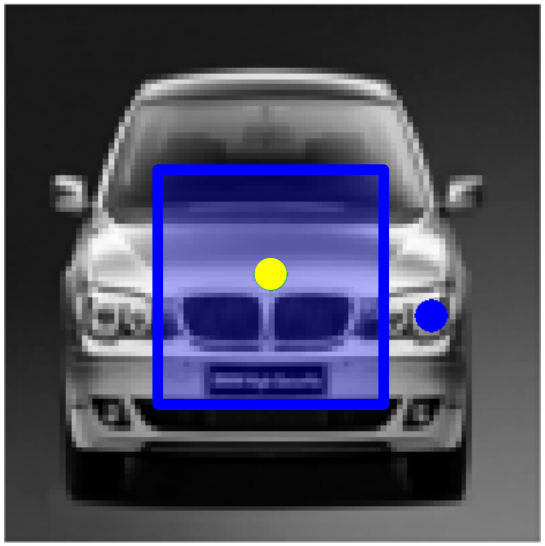
3. Heat Kernel Signature [Sun Eurographics 2009]
   - Amount of heat diffused from a point $\mathbf{p}$ over time.
   - Based on the eigenvectors φ_i and eigenvalues λ_i of $\mathbf{L}$

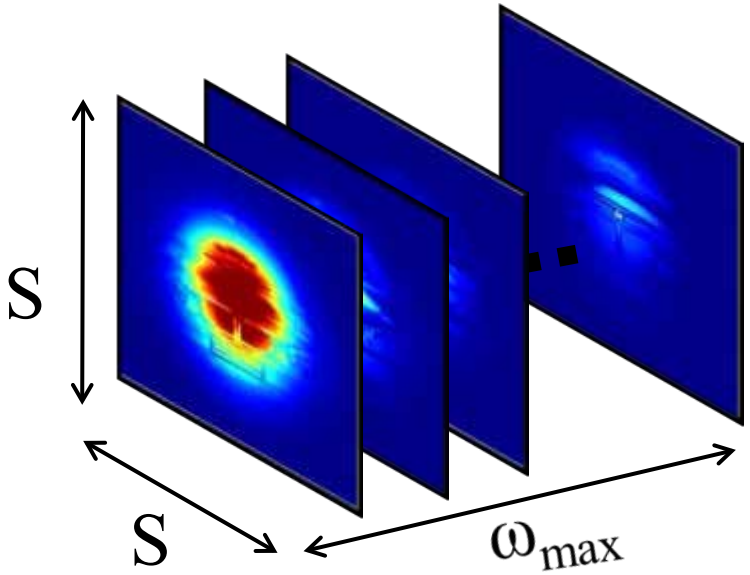
$$HKS(\mathbf{p},t) = \sum_{i=0} e^{-\lambda_i t} \varphi_i^2(\mathbf{p})$$



DaLI Computation

4. We take the whole patch to make DaLI robust to 2D pixel noise.
5. Take it in frequency domain → robustness to scale/light changes.



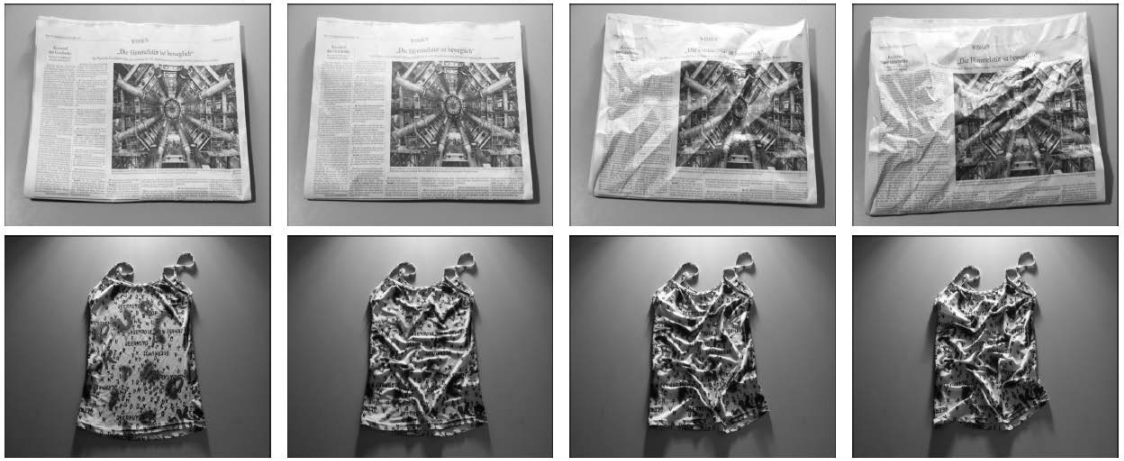


Results: dataset

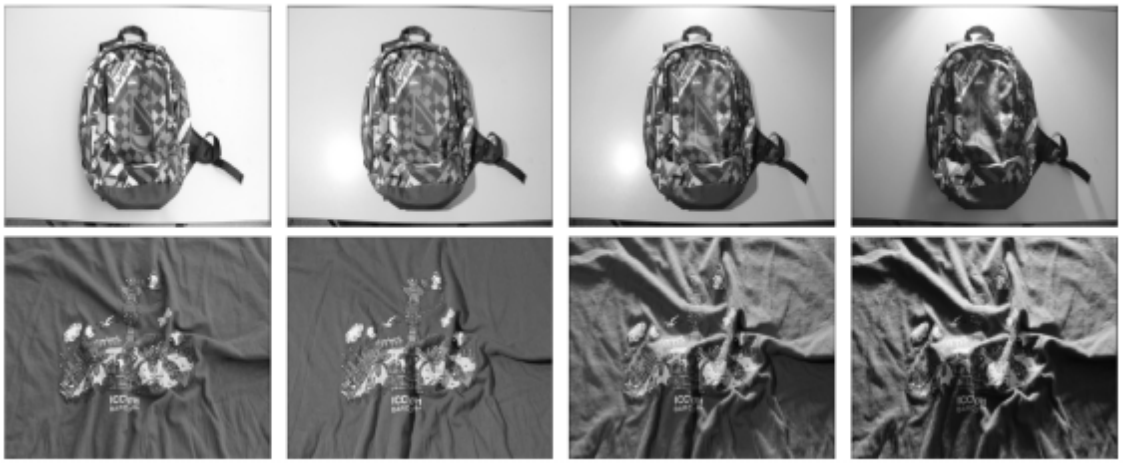
- Deformable and varying illumination dataset
  - 12 objects of different materials
  - 192 image pairs manually annotated

Results: dataset

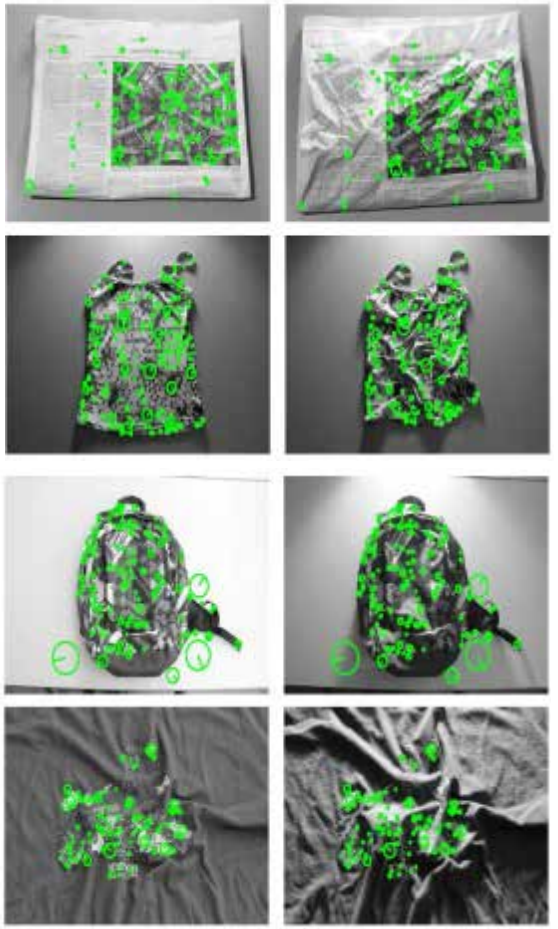
4 deformation levels per object



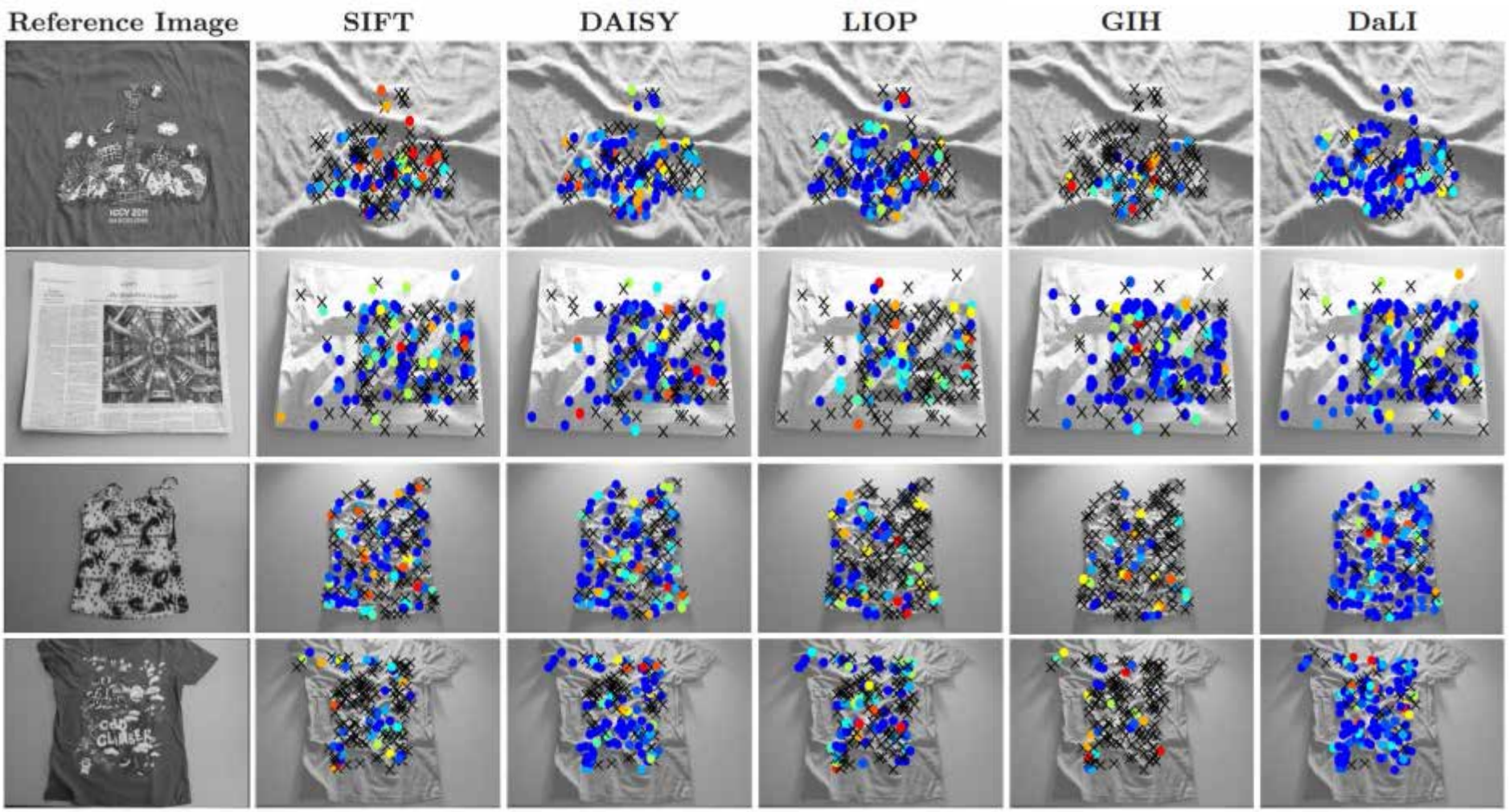
4 lighting conditions per object



Manual annotation of correspondences

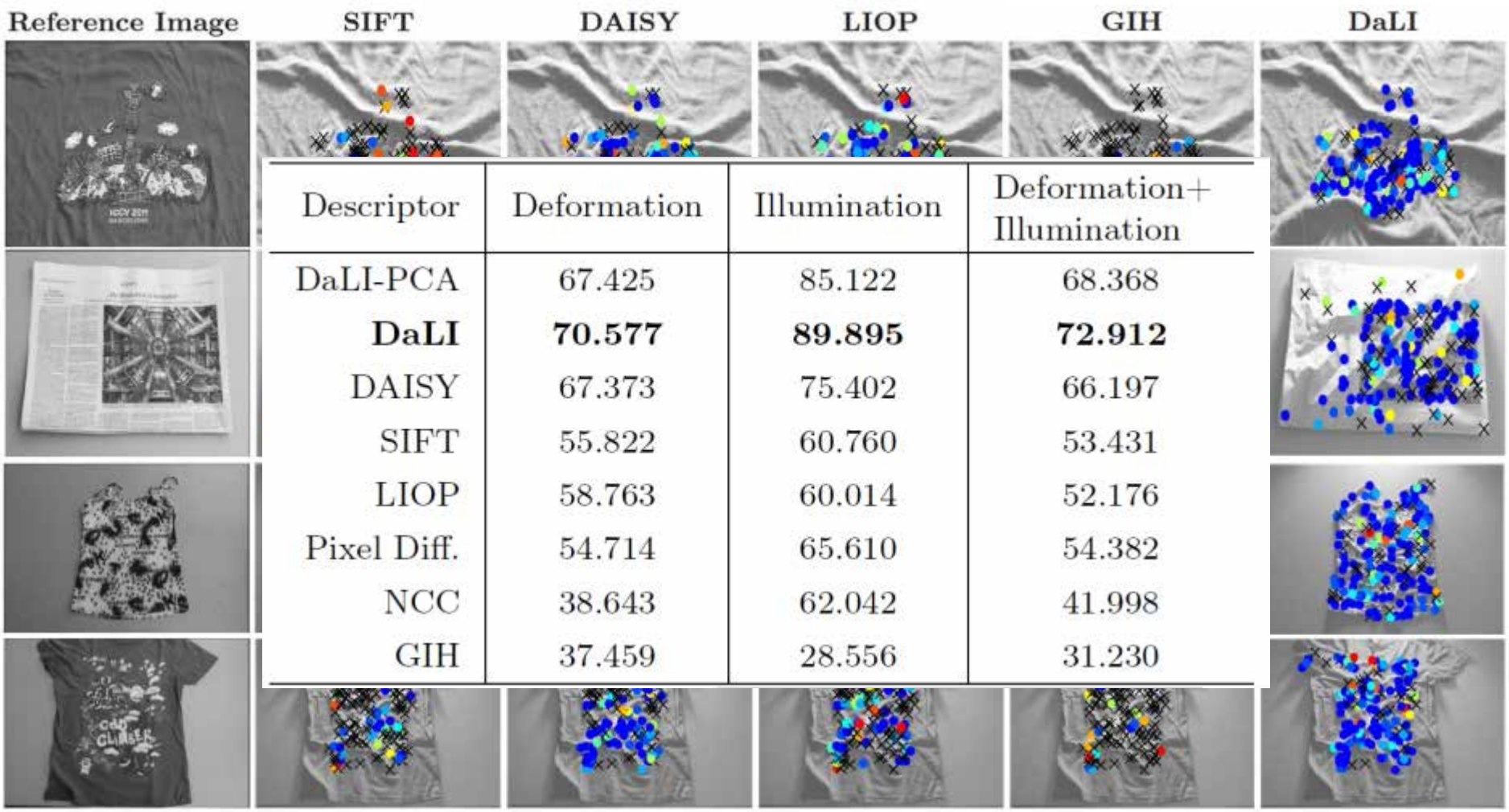


Results



[CVPR 2011]

Results

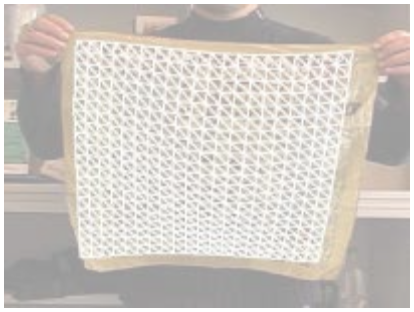
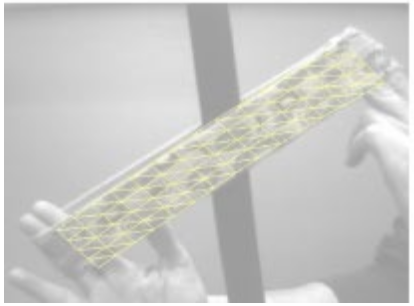


Descriptor	Deformation	Illumination	Deformation+ Illumination
DaLI-PCA	67.425	85.122	68.368
DaLI	**70.577**	**89.895**	**72.912**
DAISY	67.373	75.402	66.197
SIFT	55.822	60.760	53.431
LIOP	58.763	60.014	52.176
Pixel Diff.	54.714	65.610	54.382
NCC	38.643	62.042	41.998
GIH	37.459	28.556	31.230

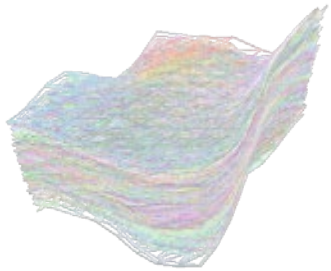
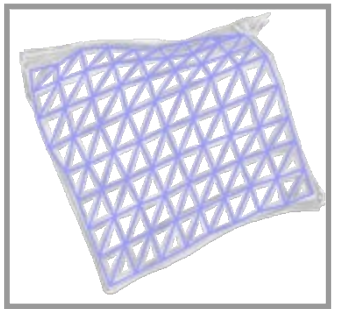
[CVPR 2011]

Outline

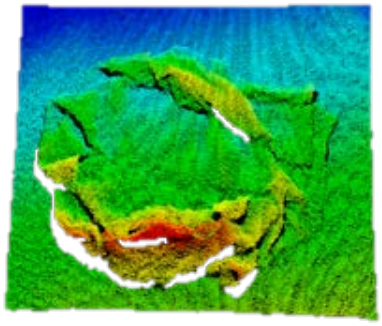
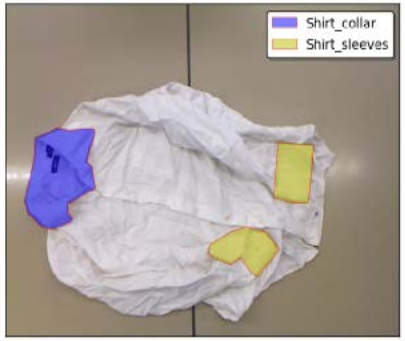
Non-Rigid Detection
(ECCV'08, CVPR'09)

Limitations of Linear Formulations
(ECCV'10, CVPR'12, PAMI'13)

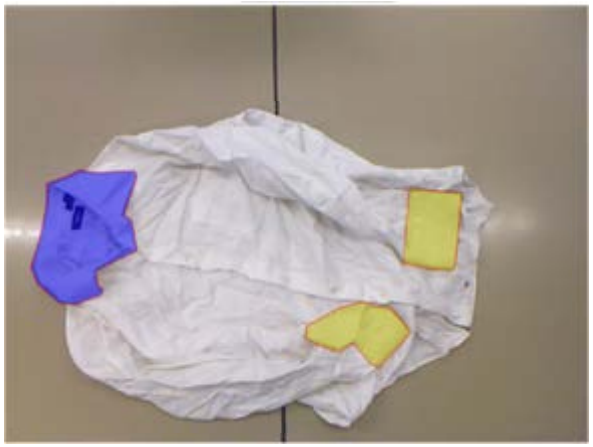



Non-Rigid Recognition
(CVPR'11, ICRA'12, IROS'13)

Visual Recognition for Manipulation

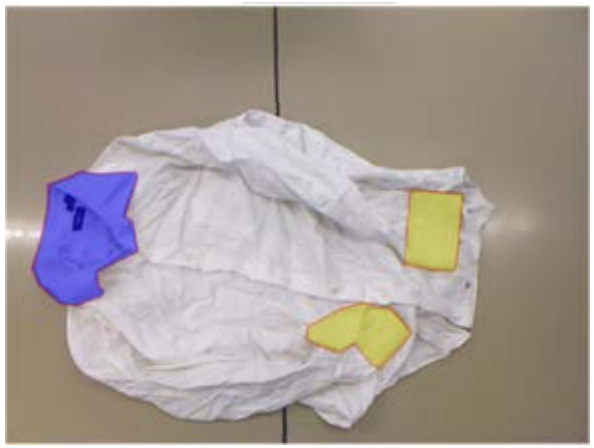
- Methods based on single RGB images cannot handle highly wrinkled clothes
- Non-rigid recognition tasks:
  - Garment recognition
  - Recognition of specific cloth parts
  - Grasping point detection



RGB

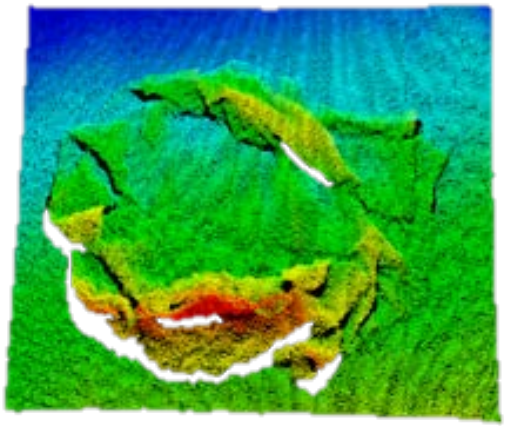
Visual Recognition for Manipulation

- Methods based on single RGB images cannot handle highly wrinkled clothes
- Non-rigid recognition tasks:
  - Garment recognition
  - Recognition of specific cloth parts
  - Grasping point detection



RGB

+



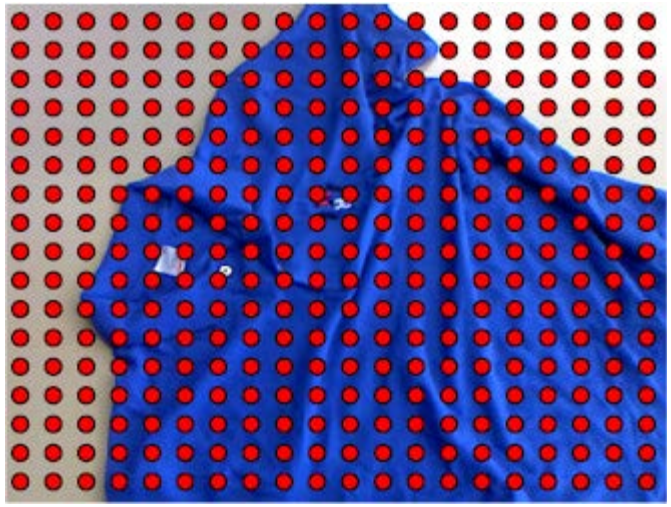
Depth

Task: collar detection

1) Part detection based on sliding window approach using a bag of visual and depth words

Task: collar detection

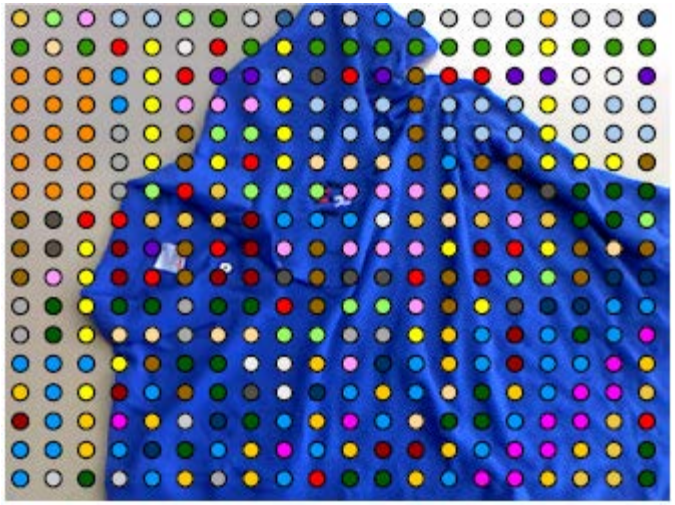
1) Part detection based on sliding window approach using a bag of visual and depth words



a) Local feature extraction

Task: collar detection

1) Part detection based on sliding window approach using a bag of visual and depth words

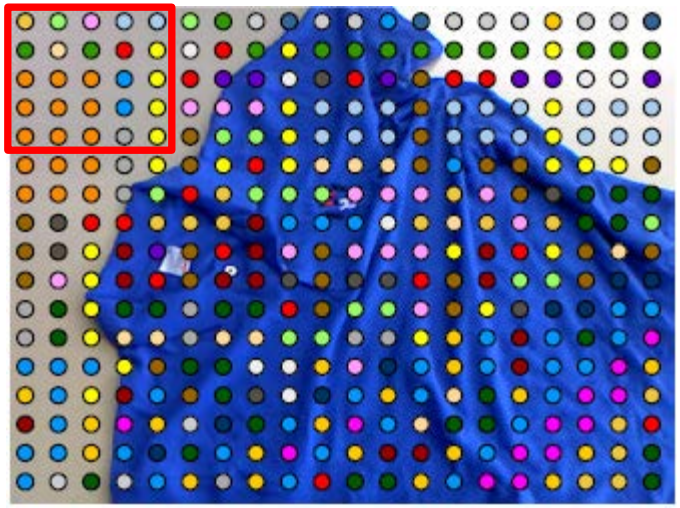


a) Local feature extraction

b) Quantize features into visual words

Task: collar detection

1) Part detection based on sliding window approach using a bag of visual and depth words



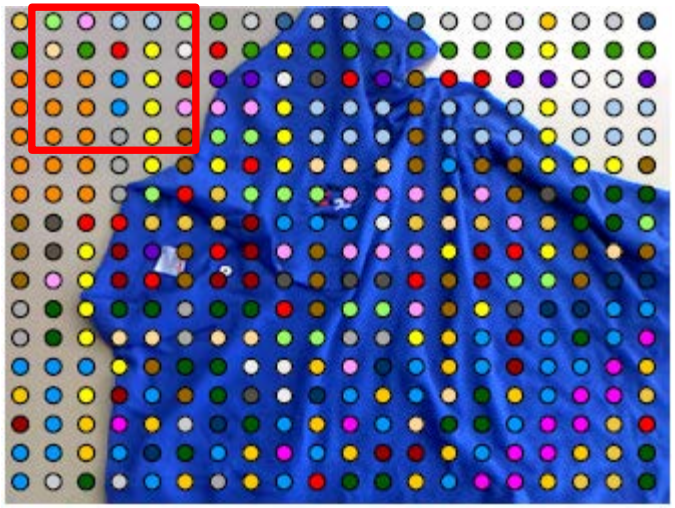
a) Local feature extraction

b) Quantize features into visual words

c) Localize part using sliding window

Task: collar detection

1) Part detection based on sliding window approach using a bag of visual and depth words



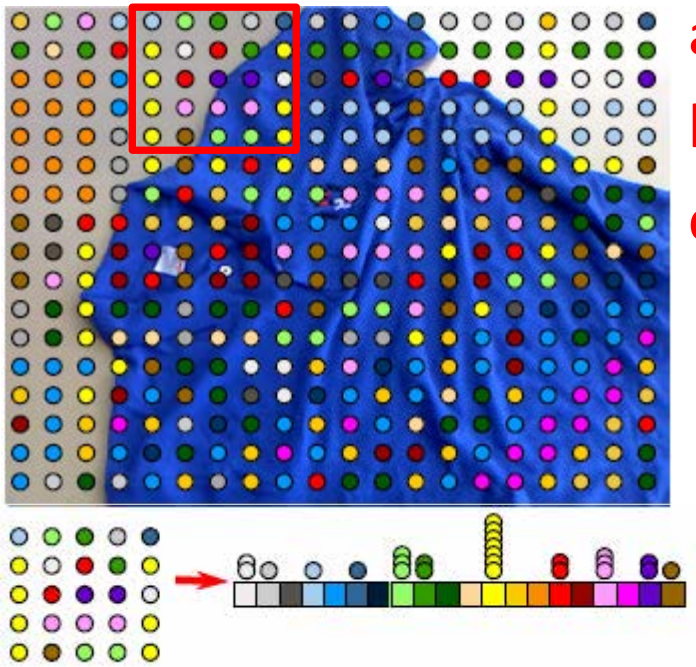
a) Local feature extraction

b) Quantize features into visual words

c) Localize part using sliding window

Task: collar detection

1) Part detection based on sliding window approach using a bag of visual and depth words



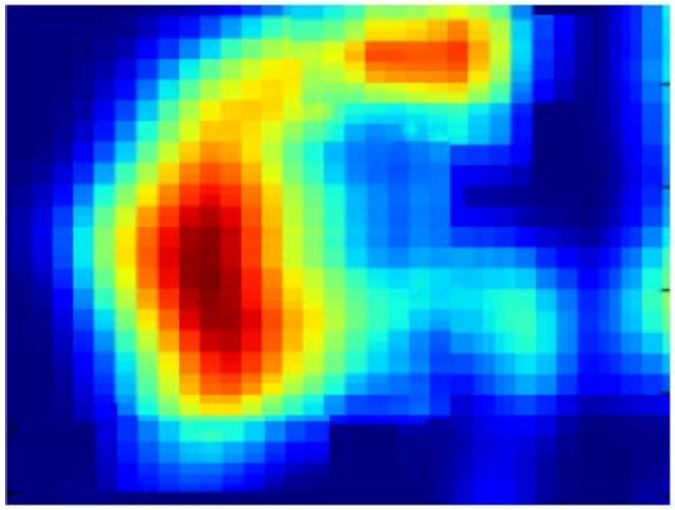
a) Local feature extraction

b) Quantize features into visual words

c) Localize part using sliding window

Task: collar detection

1) Part detection based on sliding window approach using a bag of visual and depth words



a) Local feature extraction
b) Quantize features into visual words
c) Localize part using sliding window
d) Combine responses from all windows in a probability map

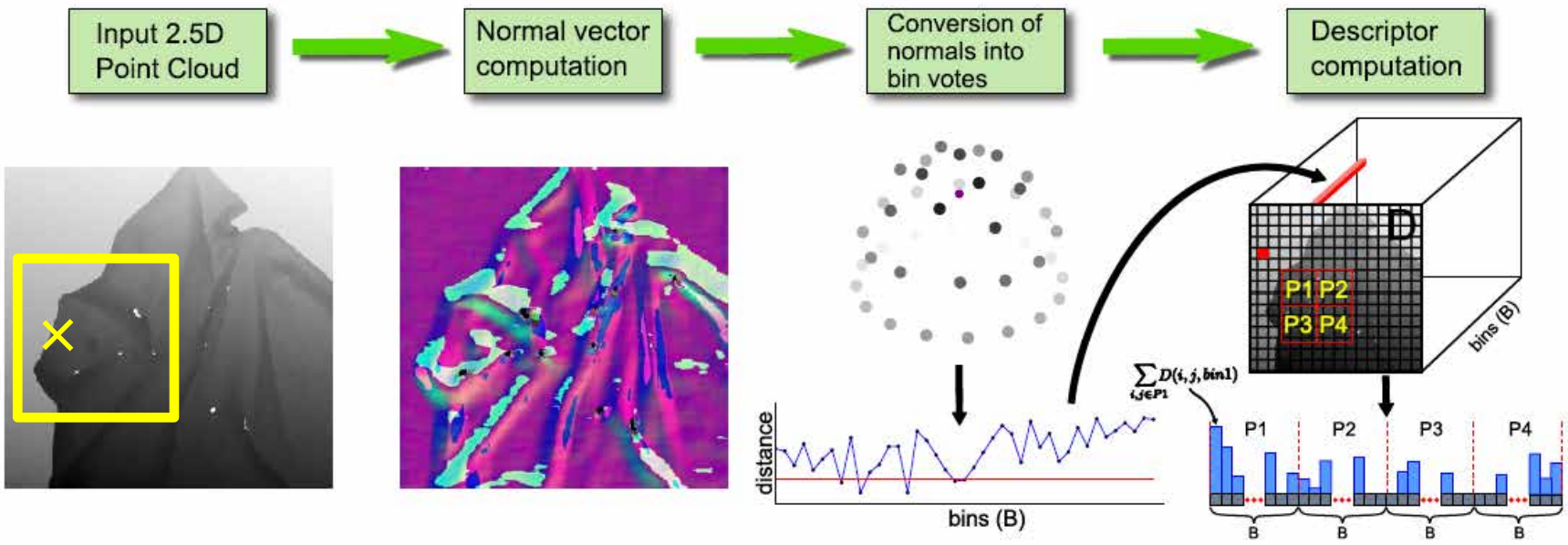
Task: collar detection

1) Part detection based on sliding window approach using a bag of visual and depth words

a) Local feature extraction
b) Quantize features into visual words
c) Localize part using sliding window
d) Combine responses from all windows in a probability map
e) Maxima are good collar candidates

Task: lapel detection

2) Detect lapel using Fast Integral Normal 3D (FINDDD) shape descriptor



Integral imaging → Allows very fast computation of the descriptor for neighboring pixels.

Collar detection > Lapel detection > Grasping

[ICRA 2012]

FINDDD Results

Garment Recognition

	RBF-χ^2 SVM — mAP		
Garment	FINDDD	SHOT	FPFH
Dress	66.8	61.9	**67.6**
Shirt	54.5	72.9	**79.7**
T-Shirt	**84.7**	70.1	76.5
Jeans	72.9	65.1	**77.9**
Polo	**96.0**	83.7	77.6
Sweater	84.6	92.1	**93.7**
Average	76.6	74.3	**78.8**

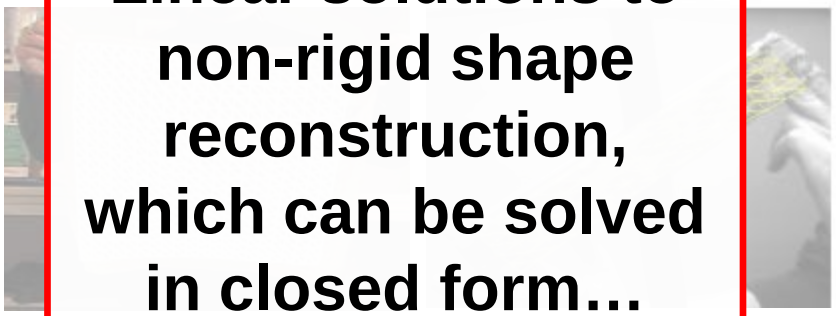
Computation Time

Descriptor	Time (s)
FINDDD (B=13)	4.0
FINDDD (B=41)	10.5
SHOT	482.5
FPFH	313.5

[IROS 2013]

Conclusions

Non-Rigid Detection



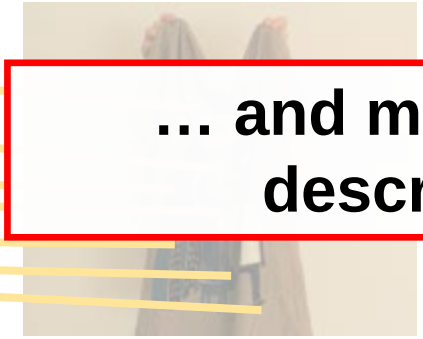
Linear solutions to non-rigid shape reconstruction, which can be solved in closed form…

Limitations of Linear Formulations

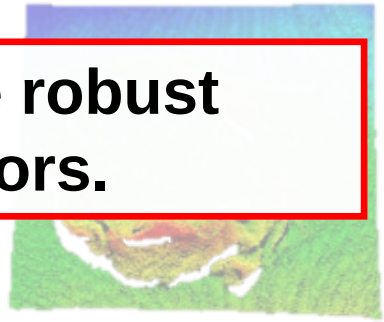
… but do not completely disambiguate the problem. This requires more sophisticated optimization approaches…

Non-Rigid Recognition

(CVPR'11, ICRA'12, IROS'13)



… and more robust descriptors.

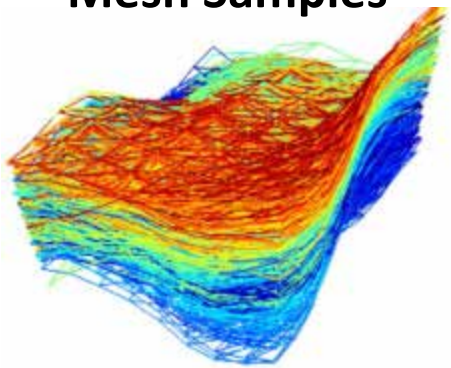


Thanks !!

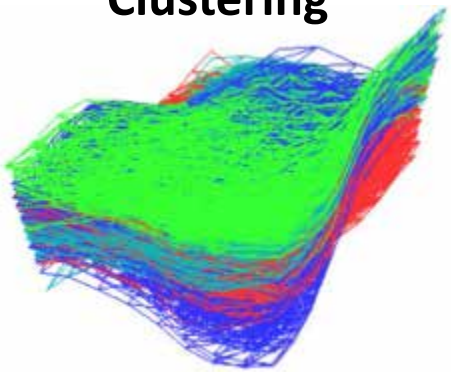
Clustering Candidate Shapes

- This yields ~10.000 shape samples (many very similar).
- Use a G-means clustering to reduce their number.
- Keep the centers of the clusters (~100-200)

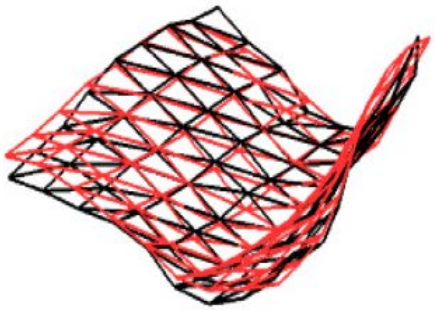
Mesh Samples

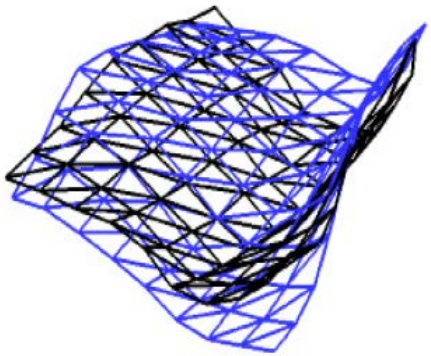


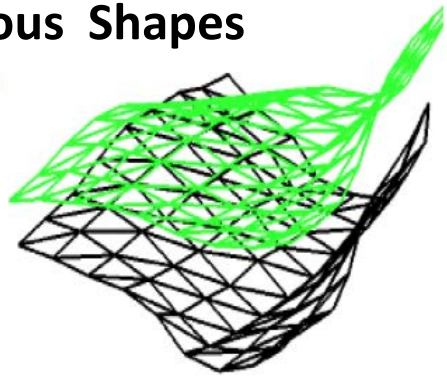
Clustering

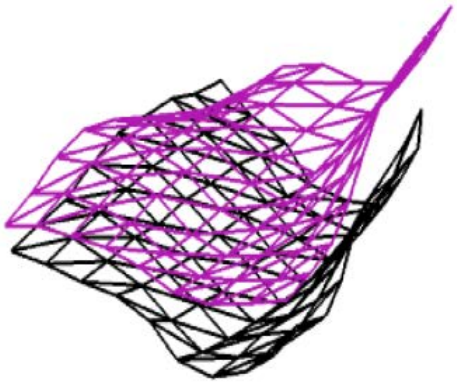


Set of Ambiguous Shapes



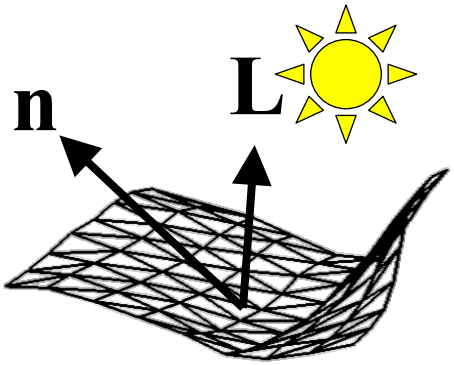






Using Shading to Disambiguate

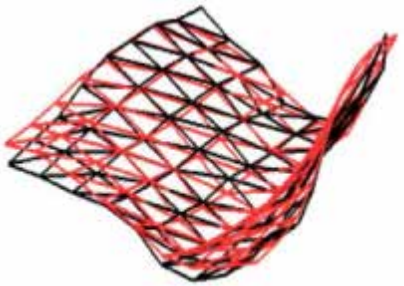
- For each candidate shape:
  1. Estimate lighting parameters

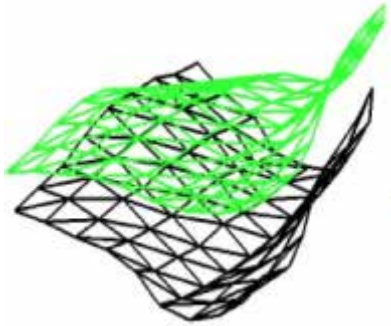


$$I_i = \rho_i \ (\mathbf{L} \cdot \mathbf{n}_i)$$

Unique unknown parameter.
Solved using least-squares.

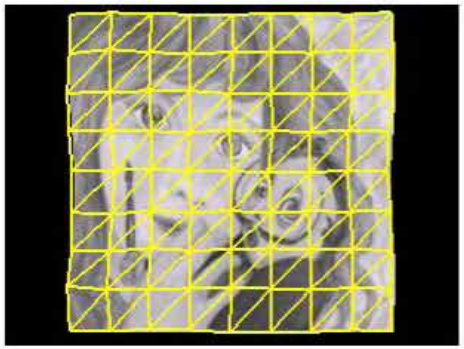
  2. Render the input image

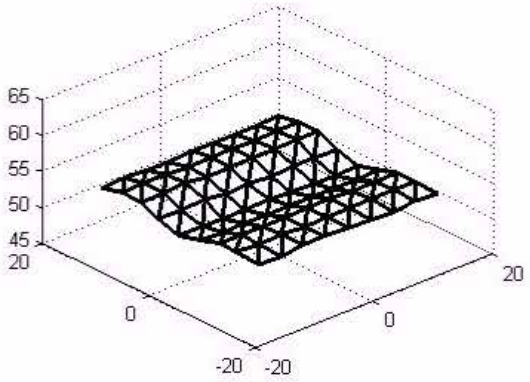


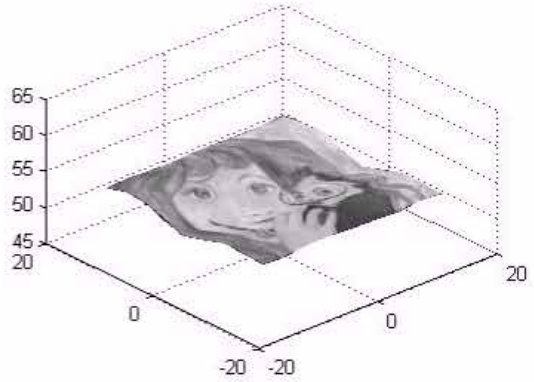


  3. Choose the candidate that best synthesizes input image

Results







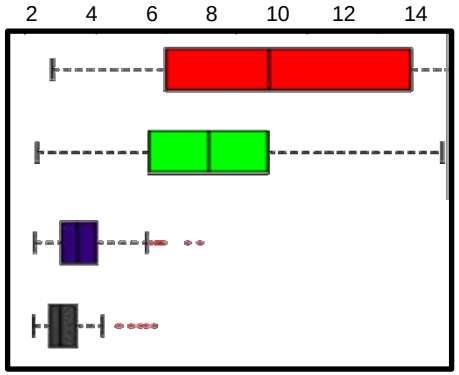
Reproj + Inext. Constraints

Reproj+Shading Constraints

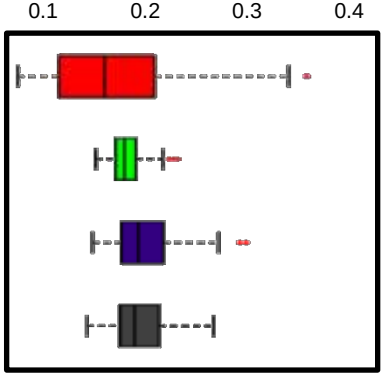
Our Approach

Best Candidate

Reconstr. Error (mm)



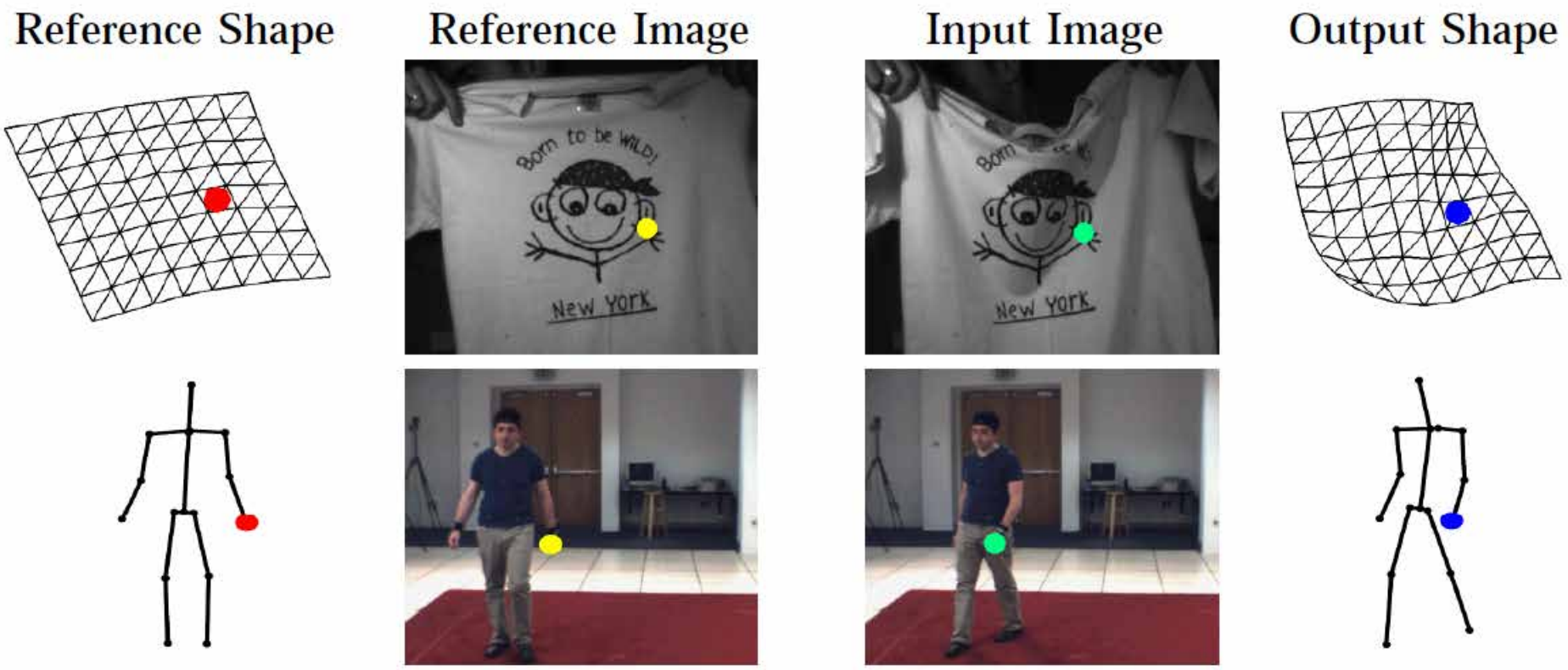
Inextens. Error (mm)



[ECCV 2010, PAMI 2013]

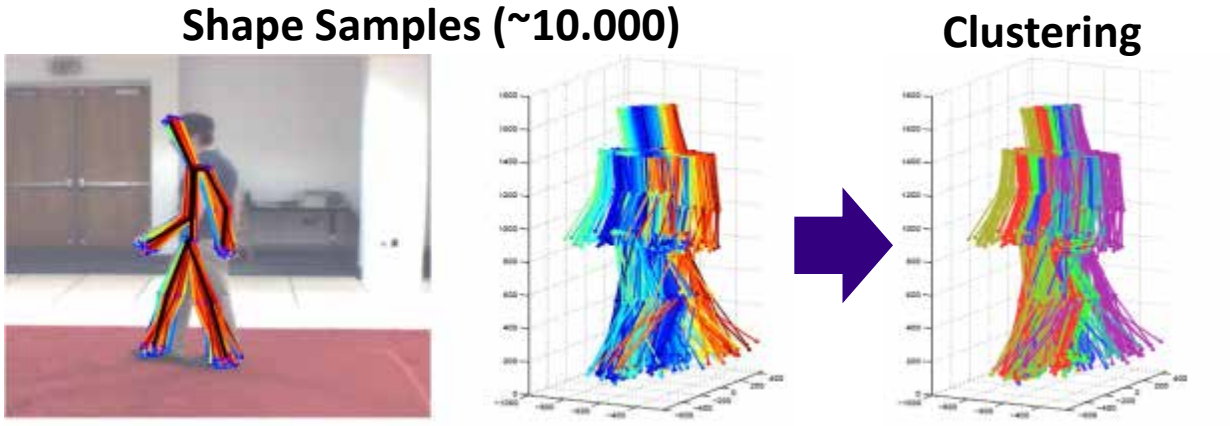
Extension to 3D Human Pose

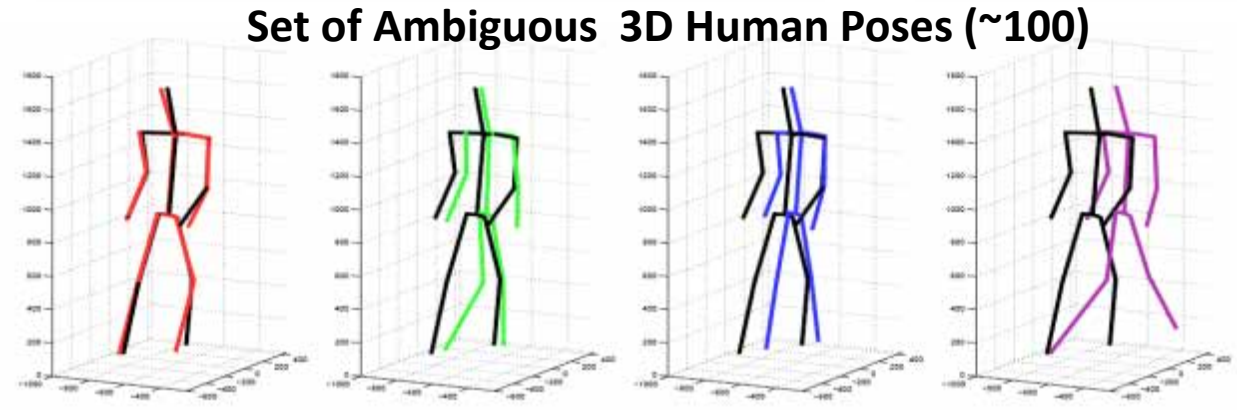
- 3D human pose and shape estimation are equivalent problems



Extension to 3D Human Pose

- Same space exploration as we did for the shapes.

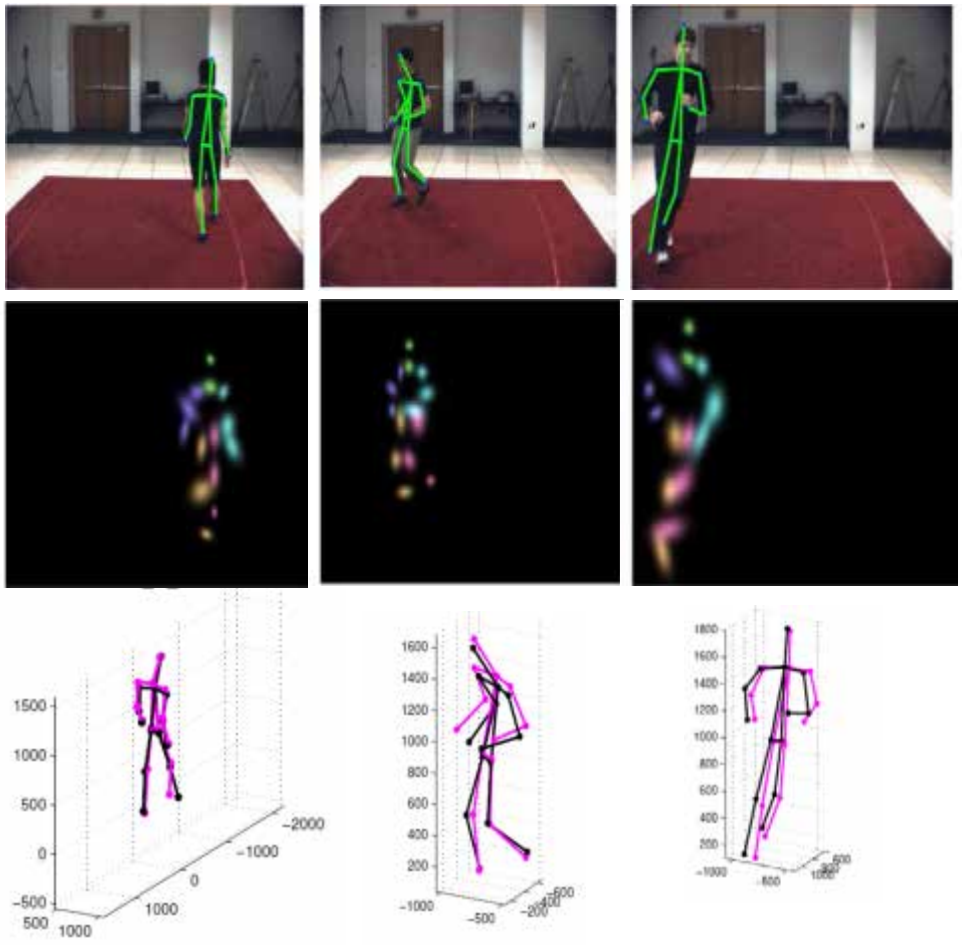




- We pick the more *human-like* candidate using a one class SVM trained with real human shapes.

Extension to 3D Human Pose

- Similar results as state-of-the-art methods that use *temporal consistency.*



	HumanEva-Walking		
	S1	S2	S3
Our Approach	99.6	108.3	127.4
Andriluka CVPR'10	-	107	-
Daubney CVPR'11	89.3	108.7	113.5

Table 1: Reconstruction Errors in mm

[CVPR 2012]

Visual Recognition for Manipulation

- Clothing RGB-D dataset with labelled parts (collar, sleeves, hood, …)