Human Segmentation, Pose Recovery and Applications

PhD thesis defense by

Meysam Madadi

Advisors: Dr. Sergio Escalera, Dr. Jordi Gonzàlez and Dr. Xavier Baró

October 13th 2017

Motivation

The final goal of computer vision is to solve problems like:

Low level

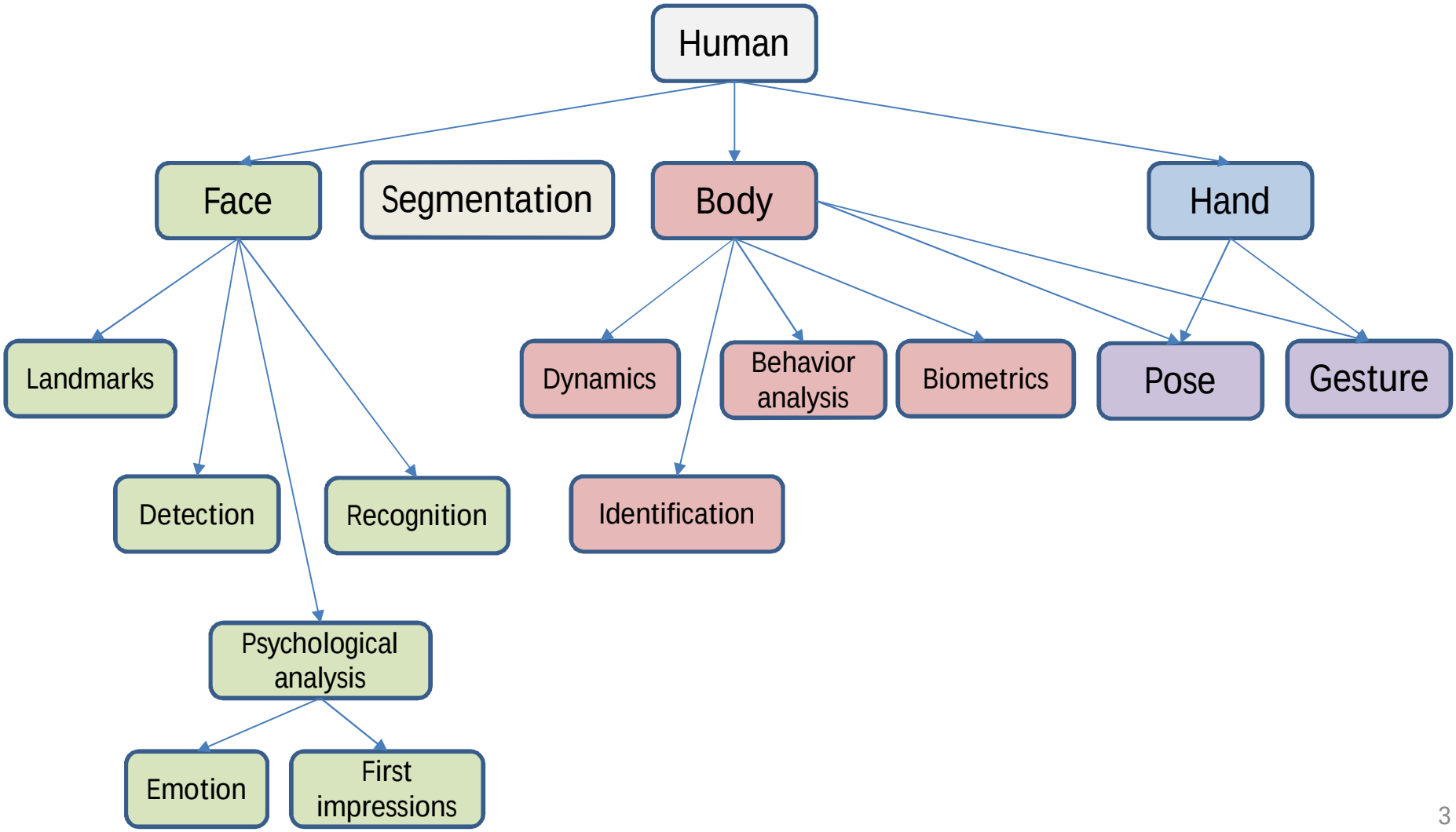
Mid level

Image | Class map | Instance map | Part map | Part map (high level)

High level

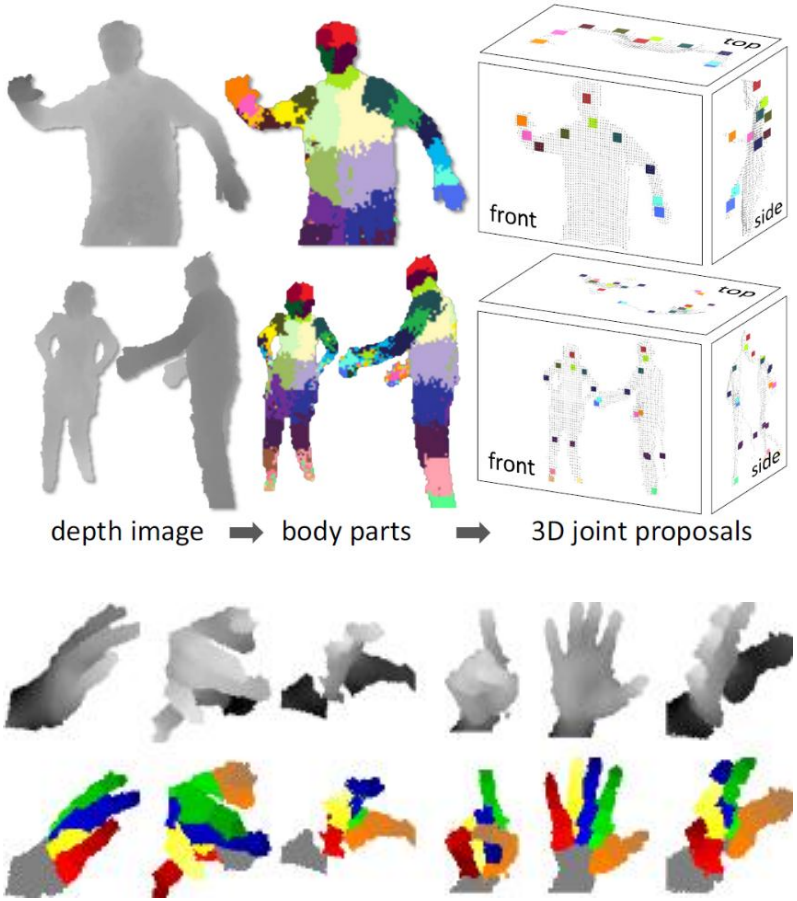
Motivation

Humans are present in most captured images and videos



Motivation

Part segmentation and pose recovery are highly correlated



[1] Shotton et al., Real-time human pose recognition in parts from single depth images. Communications of the ACM, 56(1):116–124, 2013.
[2] Vuong Le et al., Interactive facial feature localization, ECCV, 2012.

Goal

Part Segmentation
- Example-based body segmentation in depth images
- CNN-based face parsing

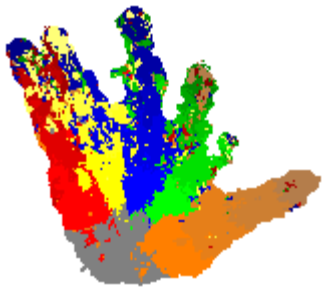
Hand pose recovery in depth images
- Top-down model fitting in a sequence of frames
- CNN-based pose regression in single frame

Applications
- Garment retexturing

Conclusions

Introduction

- Part segmentation is defined as assigning each object pixel a semantical label.
- Solutions have been proposed by both generative and discriminative methods.

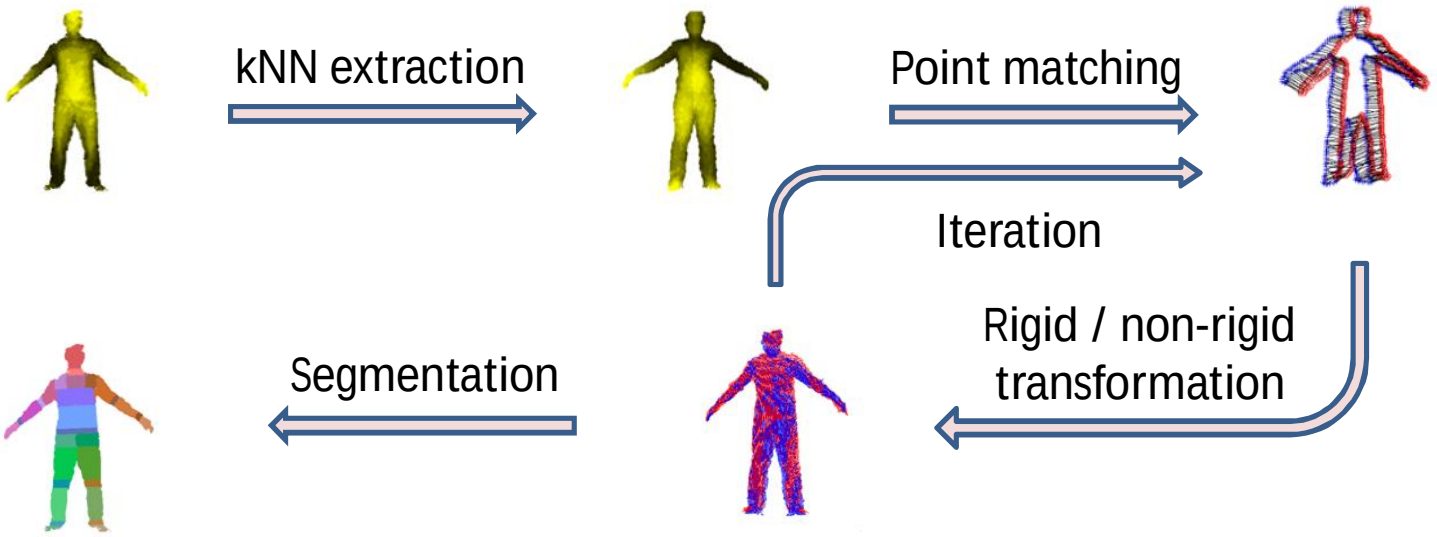


- Human body and hand have a high degree of freedom.
- Hand is a small object which can move fast.
- In the lack of data, discriminative methods may generate model drifts.

- Face is less non rigid than body. However, quite accurate segmentation is demanded for face analysis applications.
- Modeling all attributes of hair is almost intractable.

System overview

- Initial parameters of a generative model is critical,
- Objective function in generative model is minimized iteratively,
- We define segmentation as example deformation and classification.

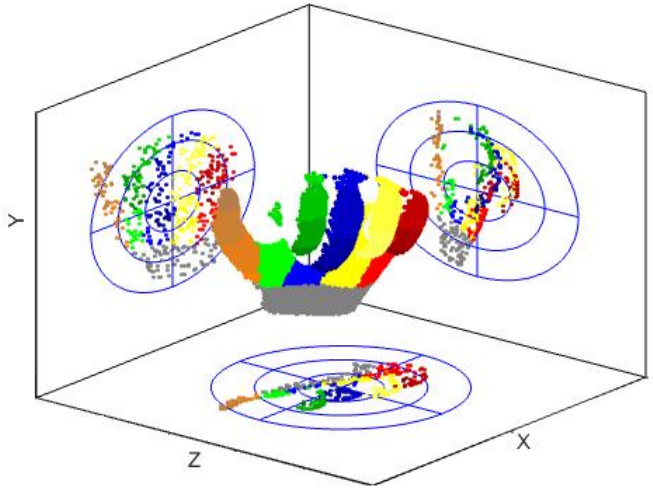


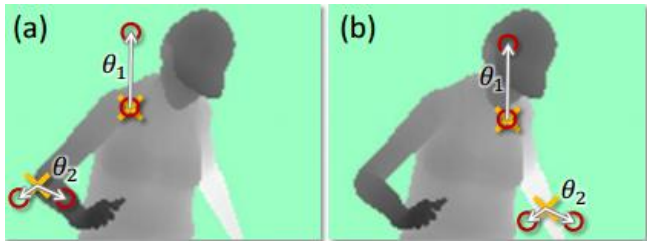
kNN extraction

- We created a shape descriptor conditioned on initial segmentation probabilities.
- Class probabilities of points are accumulated into spatial bins.

$$H_{xy}(k,c) = \sum_{i=1}^{N} \{R_{ic} | (P_i^{xy} - q^{xy}) \in bin_{xy}(k)\}$$

- Random Forest can be trained based on simple depth offset features for initial segmentation.



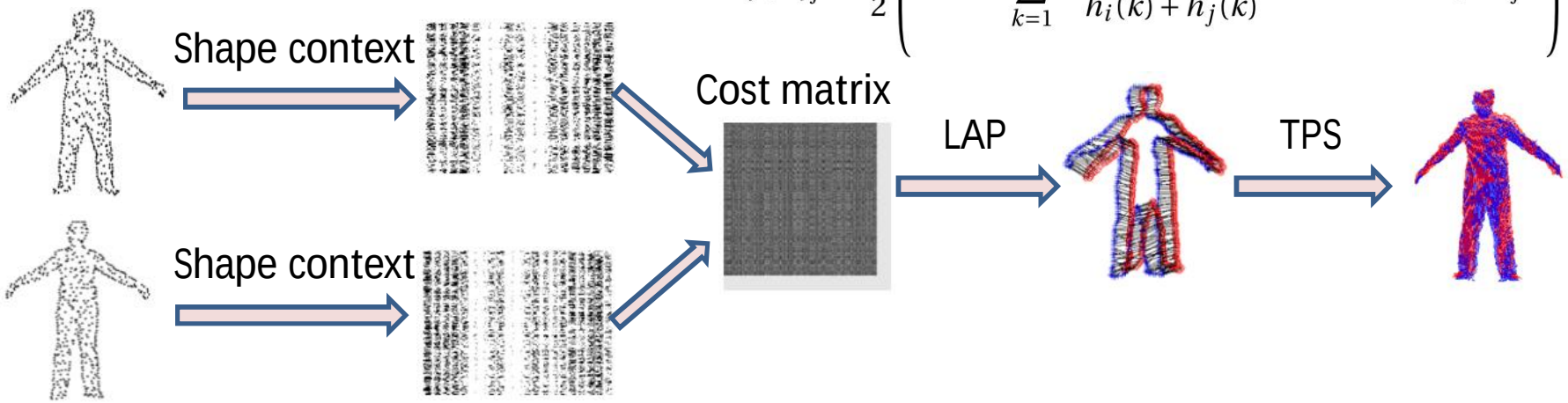


[1] Shotton et al., Real-time human pose recognition in parts from single depth images. CVPR, 2011.

Rigid / non-rigid transformation

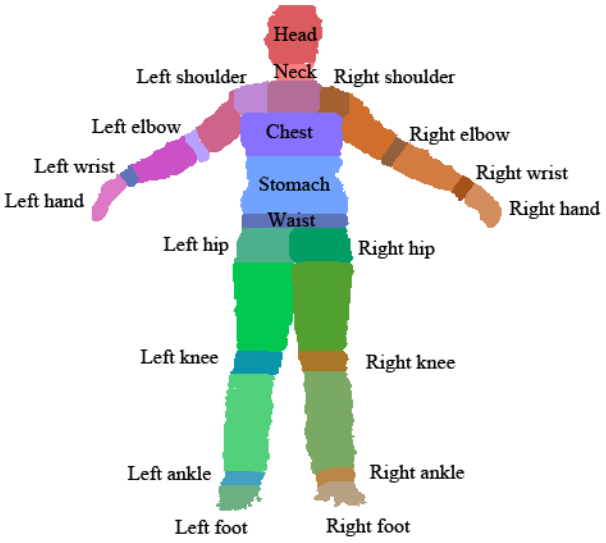
- Rigid alignment can be done when distribution of data covers all possible cases,
- Non rigid alignment can handle datasets with low amount of data,
- We define matching cost based on global and local similarity.

$$C(p_i, q_j) = \frac{1}{2}\left((1-\alpha)\sum_{k=1}^{\frac{n_r n_\theta^2}{2}} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} + \alpha(1 - \cos(t_i - t_j))\right)$$



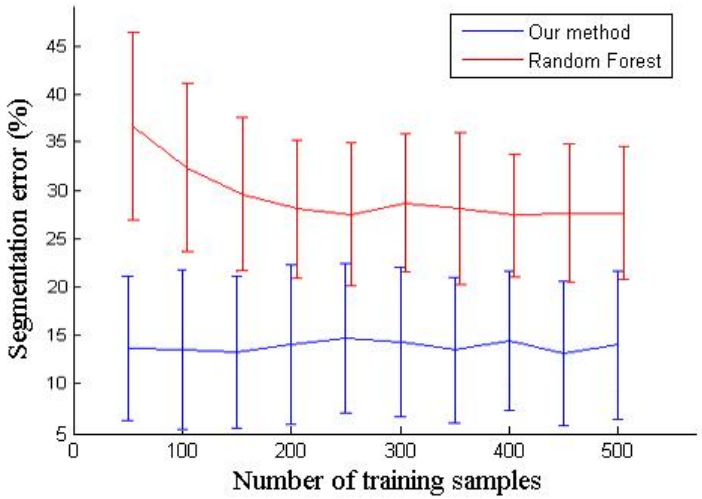
Dataset: body segmentation

- We have created a dataset of human body to evaluate our method containing of
  - RGB-D images captured by Kinect,
  - 1155 frames from 38 individuals (7 females and 31 males),
  - 29 semantical classes,
  - Front-view limb size of each person

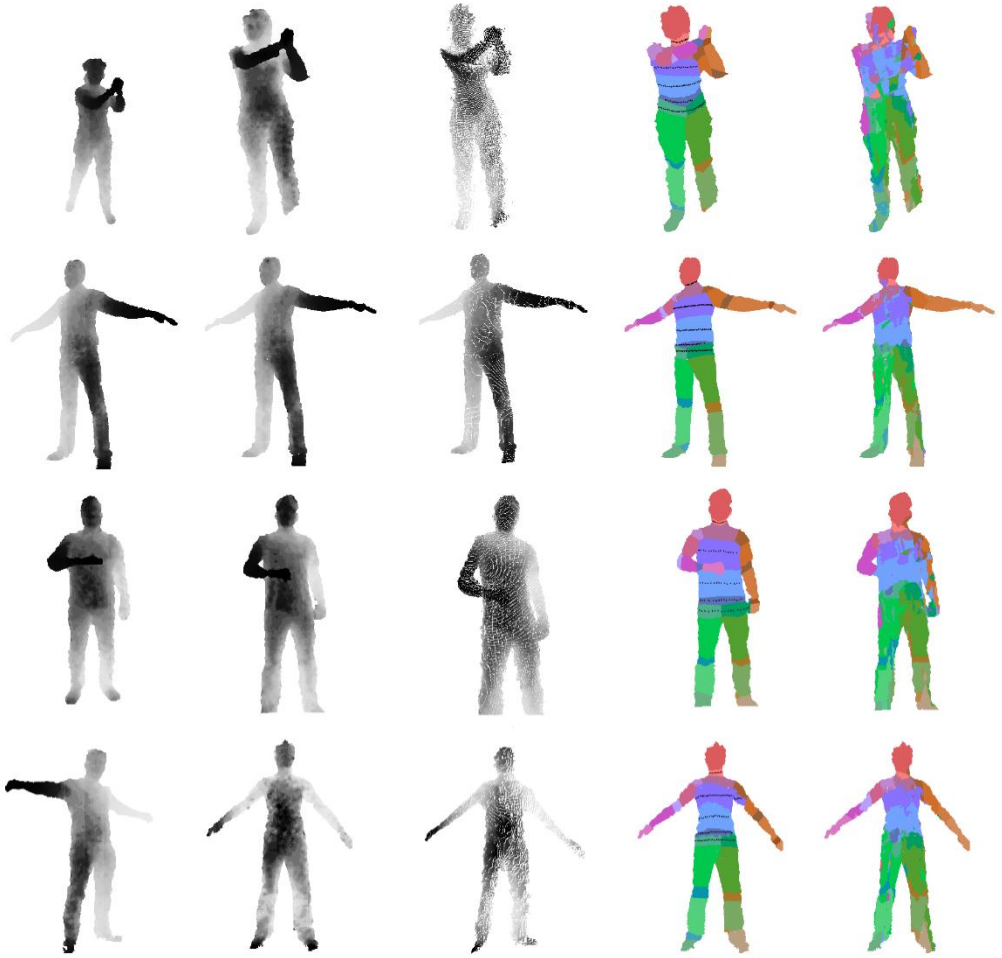


Results: body segmentation (non-rigid alignment)

Nearest neighbors are extracted based on HOG features.



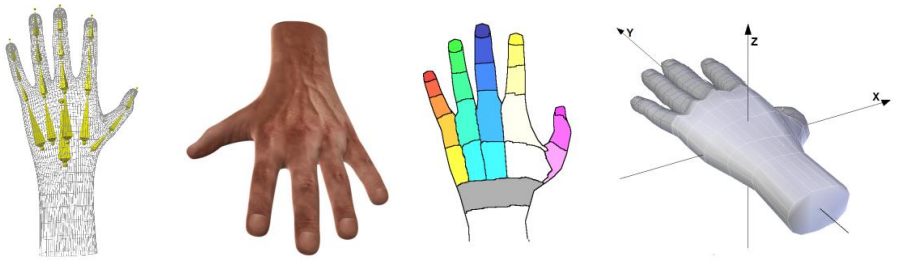
Segmentation error is the percentage of wrong classified pixels

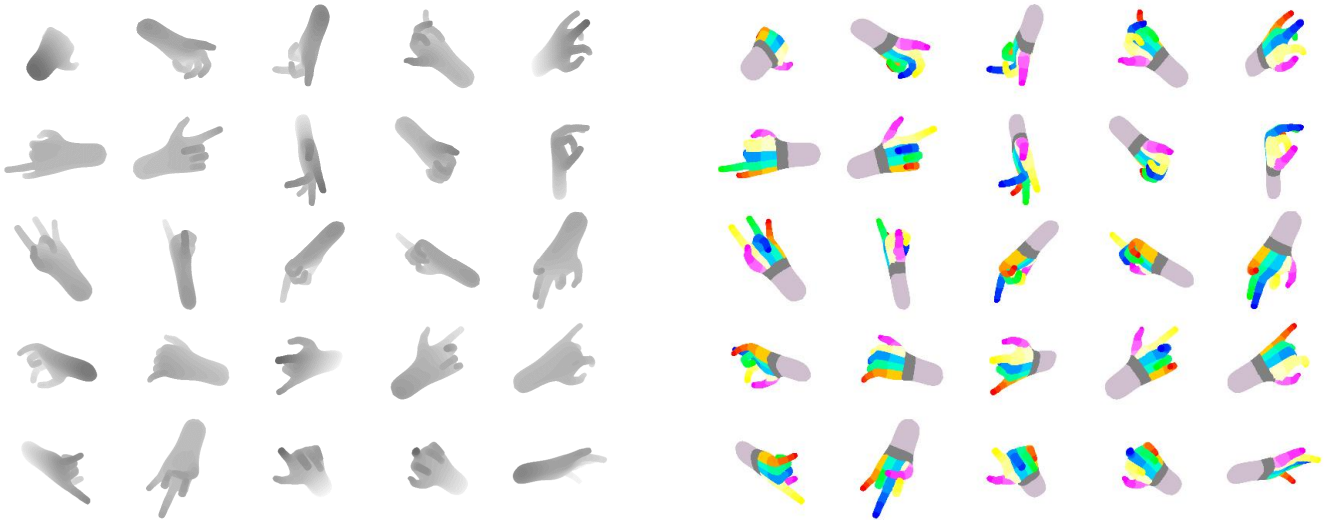


1NN | Depth image | Warped model | Ours | RF

Dataset: hand segmentation

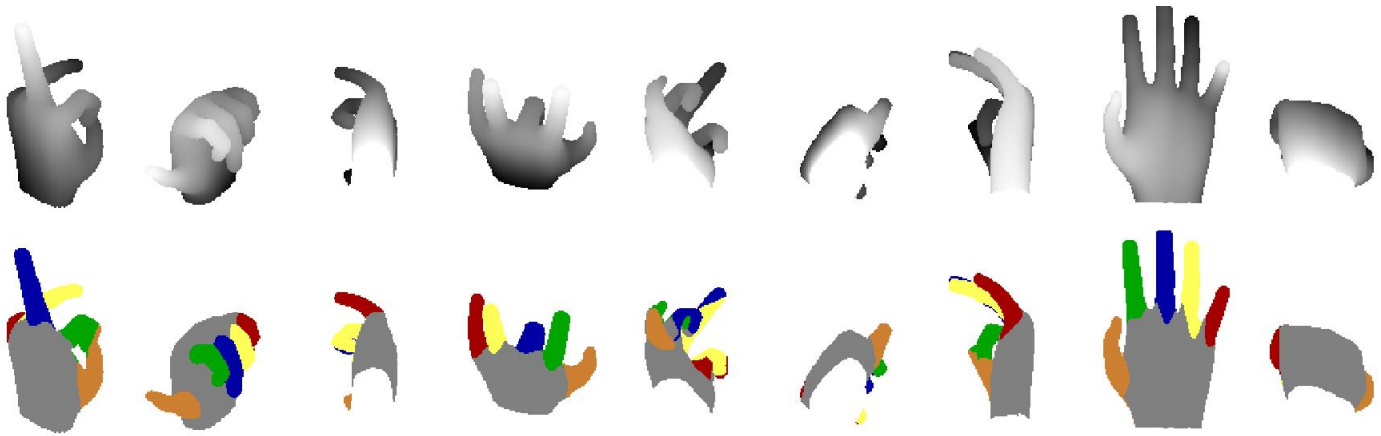
- We generated a synthetic hand dataset with natural finger movements and high degree of occlusion, consisting of
  - +600K single frames,
  - +1M sequential mocap data,
  - 25 semantical classes,
  - 20 hand joints.





Results: hand segmentation (rigid alignment)

- Results are generated based on 3NN, ICP and QDA.



Conditional random field (CRF)

- CRF is defined by Gibbs distribution as $P(\mathbf{X}=\mathbf{x}|\mathbf{I}) = \frac{1}{Z(\mathbf{I})}\exp(-E(\mathbf{x}|\mathbf{I}))$
- Gibbs energy function *E* is defined by unary and pairwise potentials terms as $E(\mathbf{x}) = \sum_{i\in\mathcal{V}} \psi_u(x_i) + \sum_{i,j\in\mathcal{E}} \psi_p(x_i, x_j)$

- Pairwise potential is defined based on compatibility function and pairwise kernels as $\psi_p(x_i, x_j) = \mu(x_i, x_j)k_{i,j}$
- Gibbs distribution can be approximated by mean field distribution in the form $Q(\mathbf{X}) = \prod_{i\in\mathcal{V}} Q_i(X_i)$ and iterative updating function

$$Q_i(x_i = l) = \frac{1}{Z_i}\exp\left(-\psi_u(x_i) - \sum_{l'\in\mathcal{L}} \mu(l, l') \sum_{i,j\in\mathcal{E}} k_{i,j} Q_j(l')\right)$$

[1] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected CRFs with Gaussian edge potentials. NIPS, 2012.

CNN-based methods

- CRF mean field approximation can be formulated by recurrent neural networks (Zheng 2015).
- Pairwise kernel can be learned based on a 4-connected graph (Liu 2015).
- Segmentation network can be trained by adversarial strategy (Luc 2016).

	adversarial training	conditional random field (ψ_u)	(ψ_p)	(end-to-end)	dilated conv.
Yu and Koltun (2015)	×	—	—	—	✓
Liu et al. (2015)	×	✓	✓	×	×
Zheng et al. (2015)	×	✓	×	✓	×
Luc et al. (2016)	✓	—	—	—	✓
Cnn (Ours)	×	—	—	—	✓
CnnGan (Ours)	✓	—	—	—	✓
CnnRnn (Ours)	×	✓	✓	✓	✓
CnnRnnGan (Ours)	✓	✓	✓	✓	✓

[1] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. CoRR, 2015.
[2] Sifei Liu et al., Multi-objective convolutional learning for face labeling. CVPR, 2015.
[3] Shuai Zheng et al., Conditional random fields as recurrent neural networks. ICCV, 2015.
[4] Pauline Luc et al., Semantic segmentation using adversarial networks. CoRR, 2016.

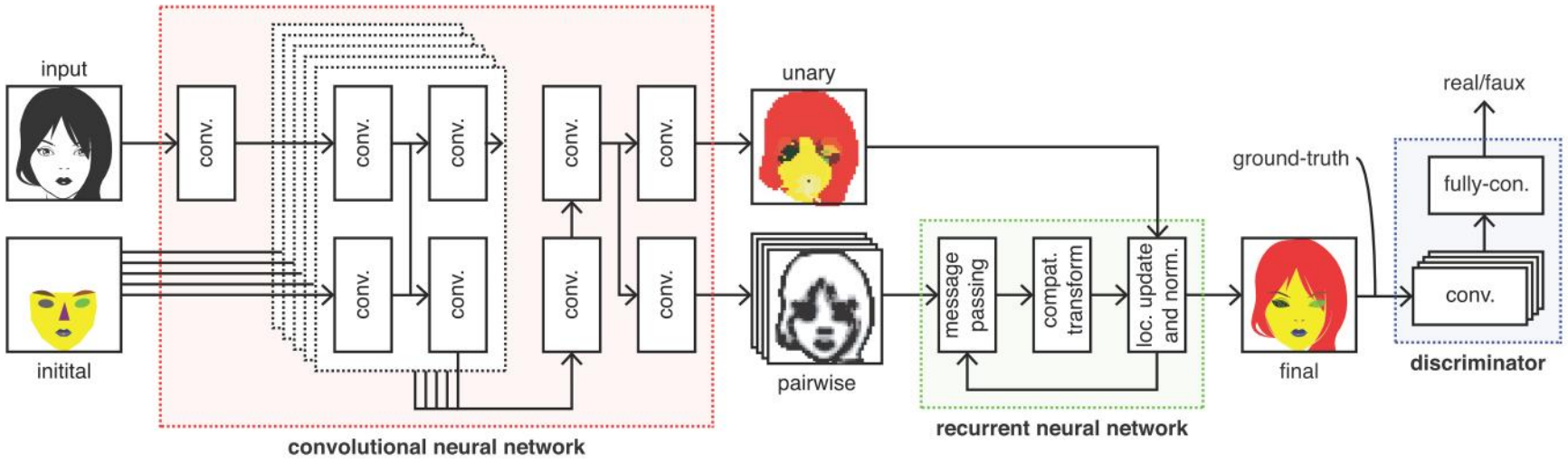
Face parsing – CNN architecture

- Network is conditioned to face landmarks,
- Pairwise kernels are learned end-to-end,
- Network is trained based on adversarial strategy,
- Discriminative network is trained based on minimax function

$$L_{dis} = -\log D_{\theta_D}\left(\mathcal{T}^{(n)}\right) - \log\left(1 - D_{\theta_D}\left(G_{\theta_G}\left(\mathcal{I}^{(n)}\right)\right)\right)$$

- Generative network is trained based on a combinatorial loss function

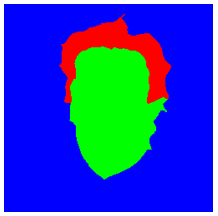
$$L_{adv} = -\log D_{\theta_D}\left(G_{\theta_G}\left(\mathcal{I}^{(n)}\right)\right) \qquad L_{seg} = -\sum_{l\in\mathcal{L}}\sum_{i\in\mathcal{V}} \mathcal{T}_{l,i}^{(n)} \log G_{\theta_G}\left(\mathcal{I}^{(n)}\right)_{l,i}$$



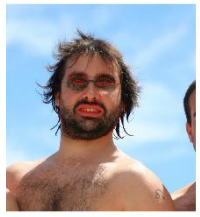
Datasets

- Parts Label dataset comprises
  - 2927 pairs of in-the-wild faces,
  - ground-truth segmentations of background, face skin (including ear skin and neck skin) and hair (including facial hair),
  - A 1500 pair training set, a 500 pair validation set and a 927 pair test set.





- Helen dataset comprises
  - 2330 pairs of in-the-wild faces,
  - ground-truth segmentations of face skin (excluding ear skin and neck skin), left eyebrow, right eyebrow, left eye, right eye, nose, upper lip, inner mouth, lower lip and hair (excluding facial hair),
  - a 2000 pair training set, a 230 pair validation set and a 100 pair test set.


[1] Andrew Kae et al., Augmenting crfs with boltzmann machine shape priors for image labeling. CVPR, 2013.
[2] Vuong Le et al., Interactive facial feature localization. ECCV, 2012.

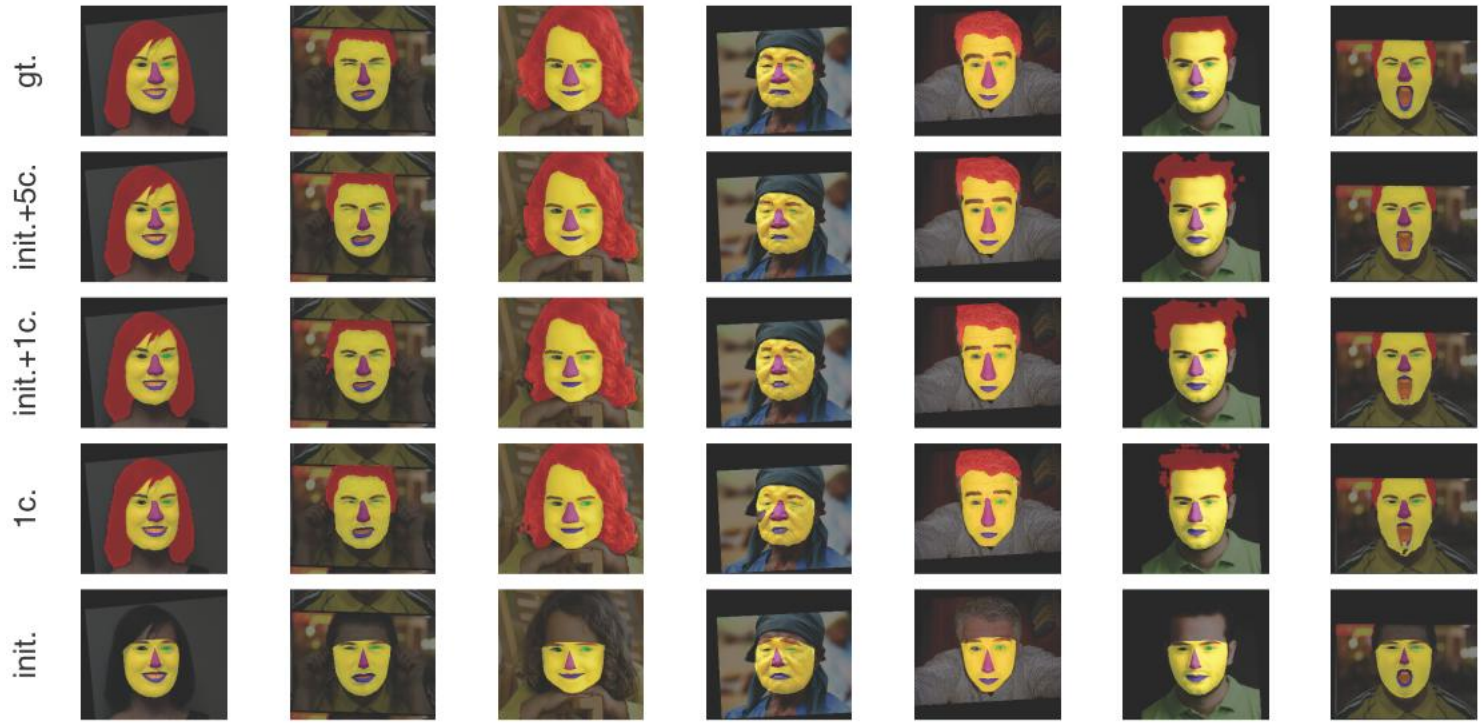
Face parsing

Results: Parts label dataset (IOU error)

				
Cnn	.9617	.9111	.7525	.8751
CnnGan	.9622	.9114	.7574	.8770
CnnRnn	**.9663**	.9177	.7795	.8878
CnnRnnGan	.9656	**.9182**	**.7808**	**.8882**

Face parsing

Results: Helen dataset (IOU error)



init.	.8253	.6358	.4855	.6527	.5325	.5568	.5757	.6001		.5405
1c.	**.9465**	.8770	.6074	.6811	.8562	.5666	.6655	.6667	**.7030**	.7300
init.+1c.	.9408	8805	.6189	.6880	.8618	.5724	.6804	.6738	.6717	.7320
init.+5c.	.9452	**.8933**	**.6987**	**.7974**	**.8884**	**.6619**	**.7467**	**.7580**	.6962	**.7873**

Results: comparing with state of the art (F1 score)

					...
Smith et. al (2013) [158]	88.20	72.20	78.50	92.20	...
Liu et. al (2015) [105]	91.20	73.40	76.80	91.20	...
Zhou et. al (2015) [214]	—	81.30	87.40	**95.00**	...
Ours	**94.36**	**82.26**	**88.73**	94.09	...

					
Smith et. al (2013) [158]	65.10	71.30	70.00	85.70	80.40
Liu et. al (2015) [105]	60.10	82.40	68.40	84.90	85.40
Zhou et. al (2015) [214]	75.40	83.60	80.90	92.60	87.30
Ours	**79.66**	**85.50**	**86.23**	**92.82**	**90.99**

					
Kae et al. (2013) [76]	—	—	—	—	94.95
Tsogkas et al. (2015) [178]	—	—	—	—	96.97
Liu et al. (2015) [105]	97.10	93.93	80.70	95.12	—
Zheng et al. (2015) [208]	—	—	—	—	96.59
Saxena et al. (2016) [146]	—	—	—	94.82	95.63
Ours	**98.25**	**95.74**	**87.69**	**96.67**	**97.16**

Introduction

- Hand pose recovery is defined as estimating 2D/3D joints locations,
- The manifold of hand pose is highly nonlinear. However, palm is rigid and has 3 DoF.
- we break the hand pose estimation problem into hierarchical optimization subtasks:
  - ❑ By using generative models in a top-down strategy while reducing the search space,
  - ❑ By using discriminative CNN regressor incorporating appearance and physical penalties.

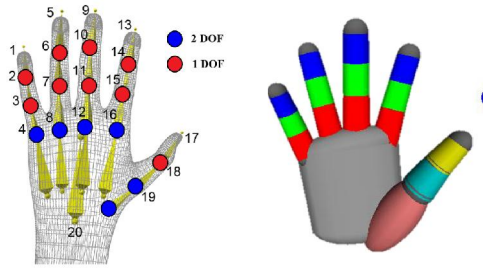
Generative top-down model

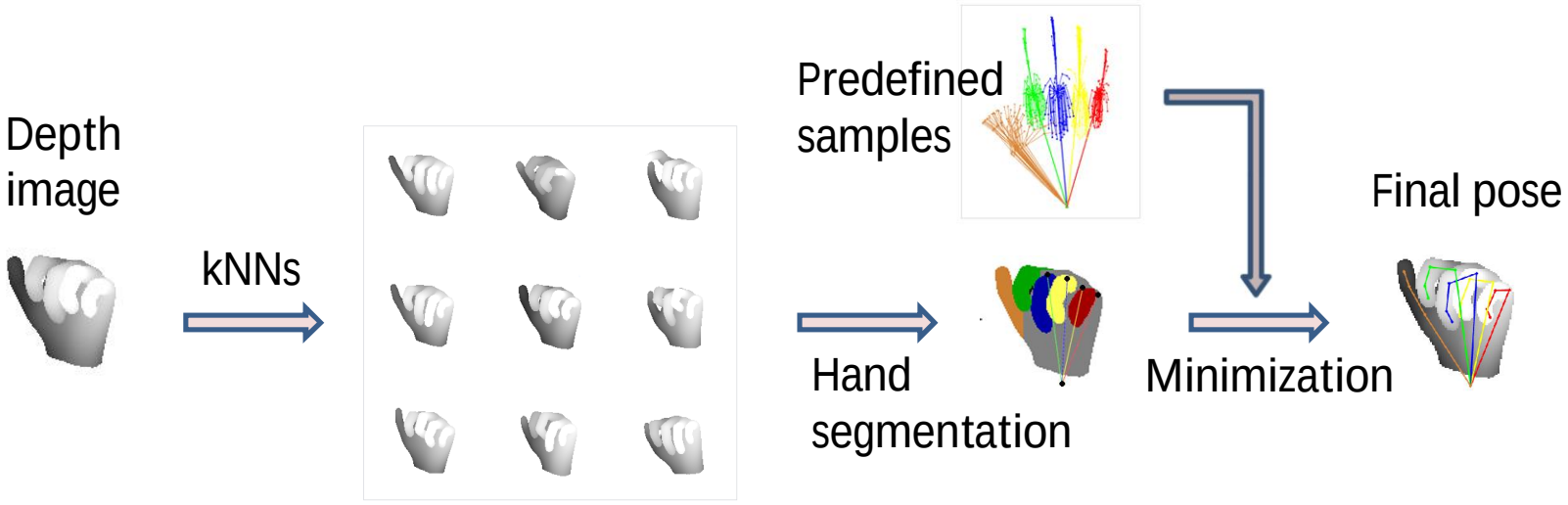
- Initializing model parameters based on similar samples (Sharp 2015),
- Separating palm and fingers regression in a hierarchical cascading model (Sun 2015),
- Advancing model fitting by enhanced objective function (Qian 2014),
- Spatial and temporal statistical model fitting (Zhou 2014).

[1] Sharp et al., Accurate, robust, and flexible real-time hand tracking. In ACM Human Factors in Computing Systems, 2015.
[2] Sun et al,. Cascaded hand pose regression. In CVPR, 2015.
[3] Qian et al,. Realtime and robust hand tracking from depth. CVPR, 2014.
[4] Zhou and F. D. la Torre. Spatio-temporal matching for human detection in video. ECCV, 2014.

Single frame pose recovery

- A set of candidate fingers are selected given:
  1. Hand segments and palm joints,
  2. A predefined set of sample fingers,
  3. A set of simple rules:
     - Joints must not be located outside the hand mask,
     - A joint must not have a depth lower than the hand surface.
- A discrepancy function *E* is minimized using a hand model. $E(h, I) = w_1 E_1 + w_2 E_2 + w_3 E_3$



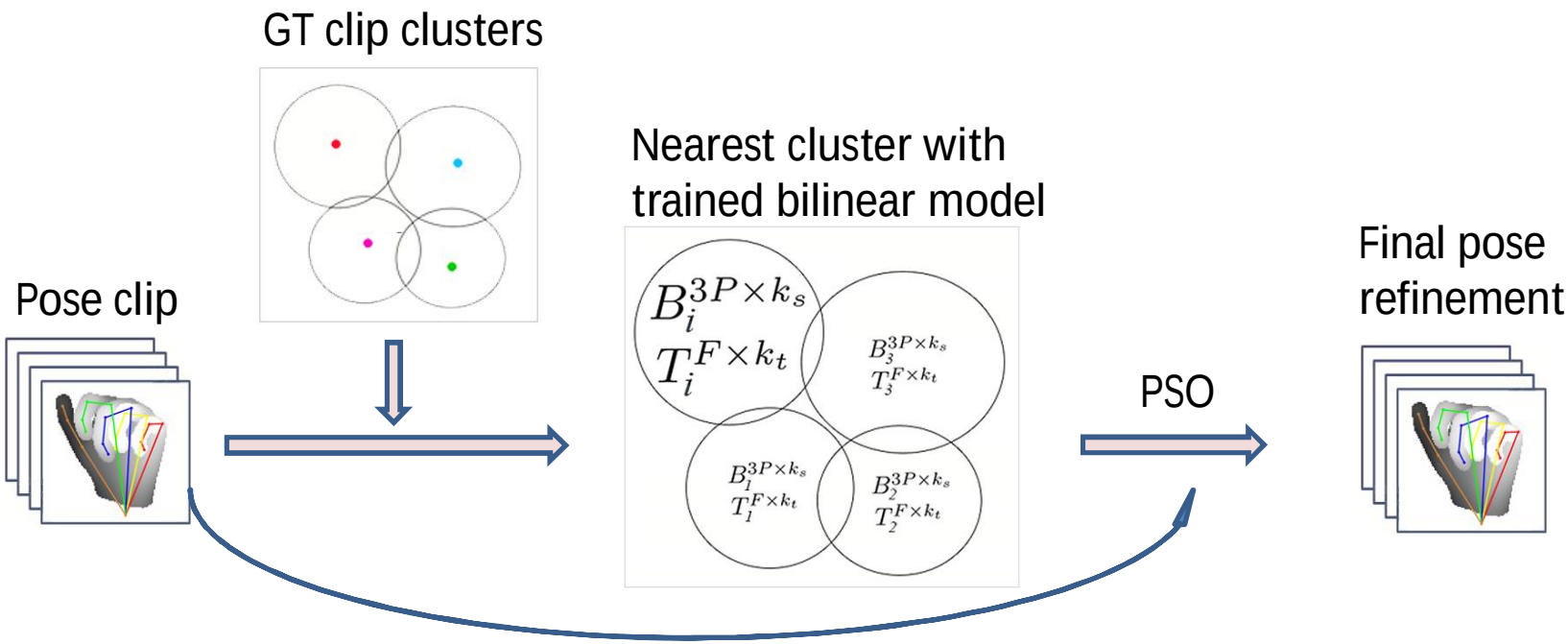


Generative top-down model

Temporal pose refinement

- A clip of last frames *Q* can be factorized through $Q = TCB^T$
- We define an objective function as

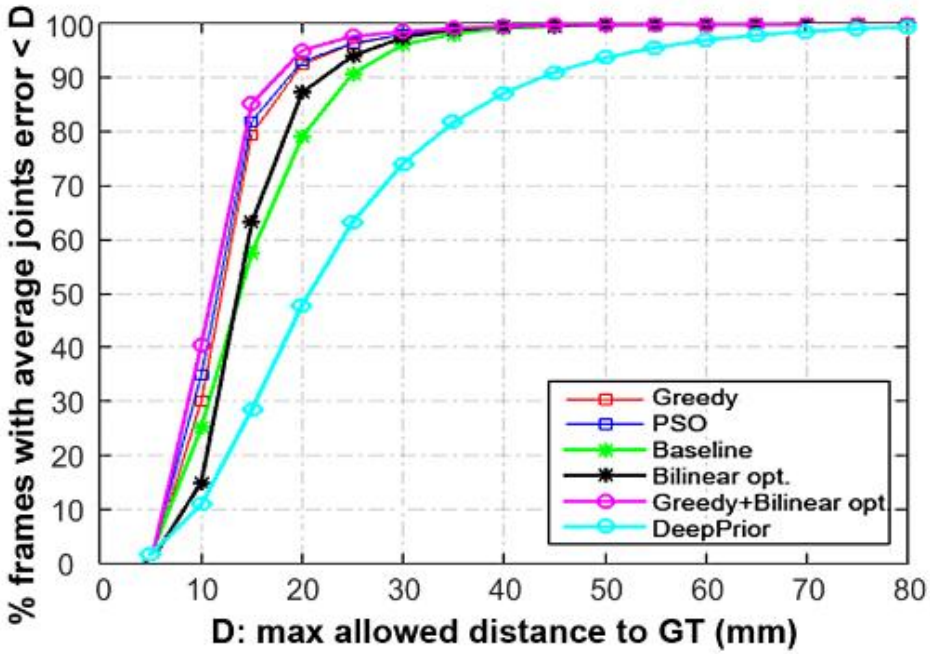
$$\underset{C}{argmin} \sum_{f=1}^{F} \sum_{i=1}^{5D} V_{fi} \left| Q_{fi} - [TCB^T]_{fi} \right| + \beta \sum_{f=1}^{F-1} \Psi^{f,f+1}$$

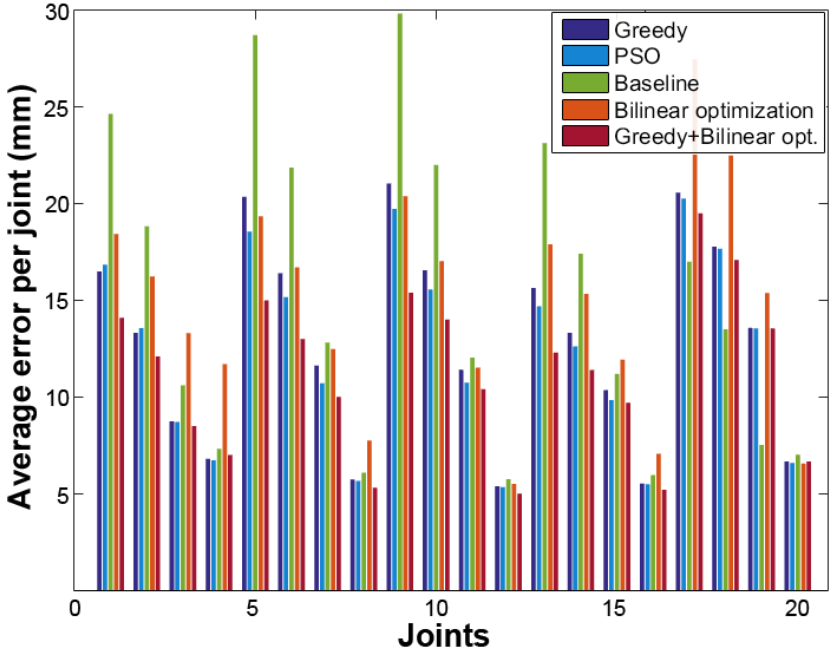


[1] I. Akhter, T. Simon, S. Khan, I. Matthews, and Y. Sheikh. Bilinear spatiotemporal basis models. TOG, 31(17), 2012.

Generative top-down model

Results: Synthetic dataset





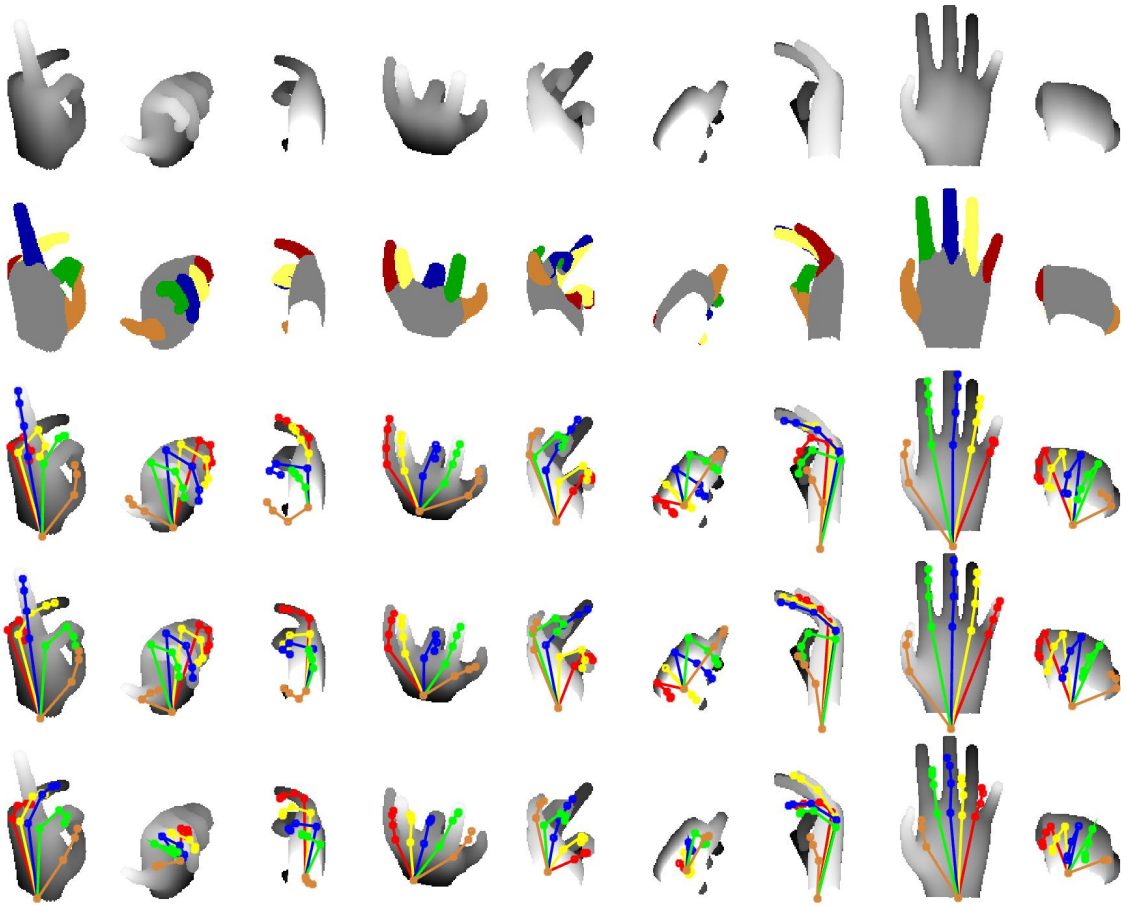
[1] Markus Oberweger et al., Hands deep in deep learning for hand pose estimation. Computer Vision Winter Workshop, 2015.

Results: Synthetic dataset

1NN

Greedy

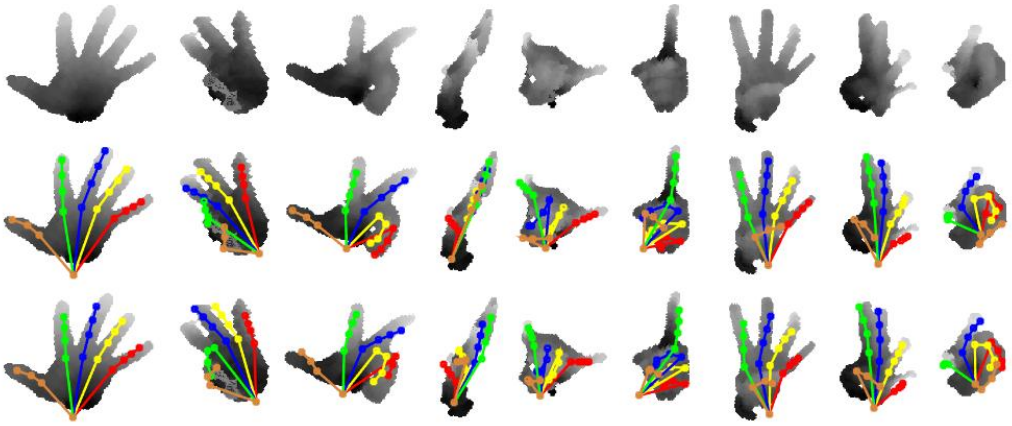
DeepPrior

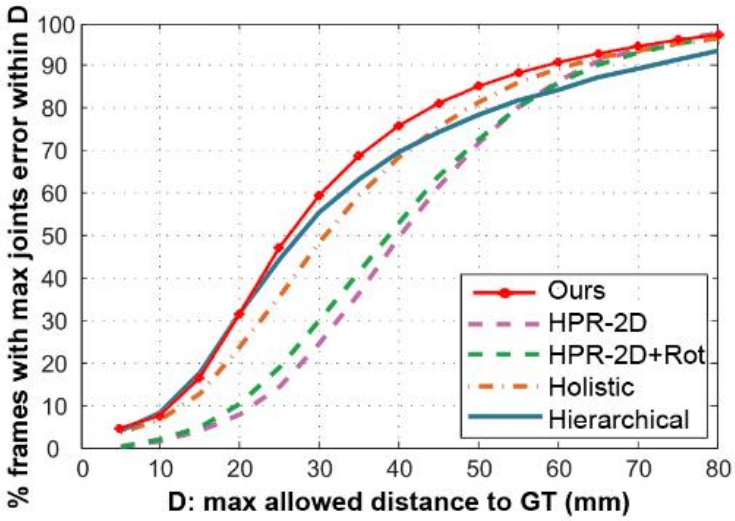


[1] Markus Oberweger et al., Hands deep in deep learning for hand pose estimation. Computer Vision Winter Workshop, 2015.

Generative top-down model

Results: MSRA dataset





Euclidean error (lower is better)

	IndexR	IndexT	MiddleR	MiddleT	RingR	RingT	LittleR	LittleT	ThumbT	Mean
Oikonomidis et al.	31.0	56.0	32.9	56.0	32.9	49.3	35.1	53.7	22.2	38.2
Choi et al.	22.6	43.5	24.0	44.9	23.1	43.1	21.8	39.5	31.1	29.8
Ge et al.	11.5	16.0	9.0	15.6	9.9	15.1	13.2	16.0	16.7	13.0
Ours (KNN+ICP)	9.5	17.3	7.7	17.1	8.3	15.5	10.6	17.7	14.8	12.8

[1] X. Sun, Y. Wei, S. Liang, X. Tang, and J. Sun. Cascaded hand pose regression. In CVPR, 2015.
[2] Iason Oikonomidis et al., Efficient model-based 3d tracking of hand articulations using kinect. BMVC, 2011.
[3] Chiho Choi et al., A collaborative filtering approach to real-time hand pose estimation. ICCV, 2015.
[4] Ge et al., Robust 3d hand pose estimation in single depth images: from single-view cnn to multi-view cnns. CVPR, 2016.

CNN-based regressor

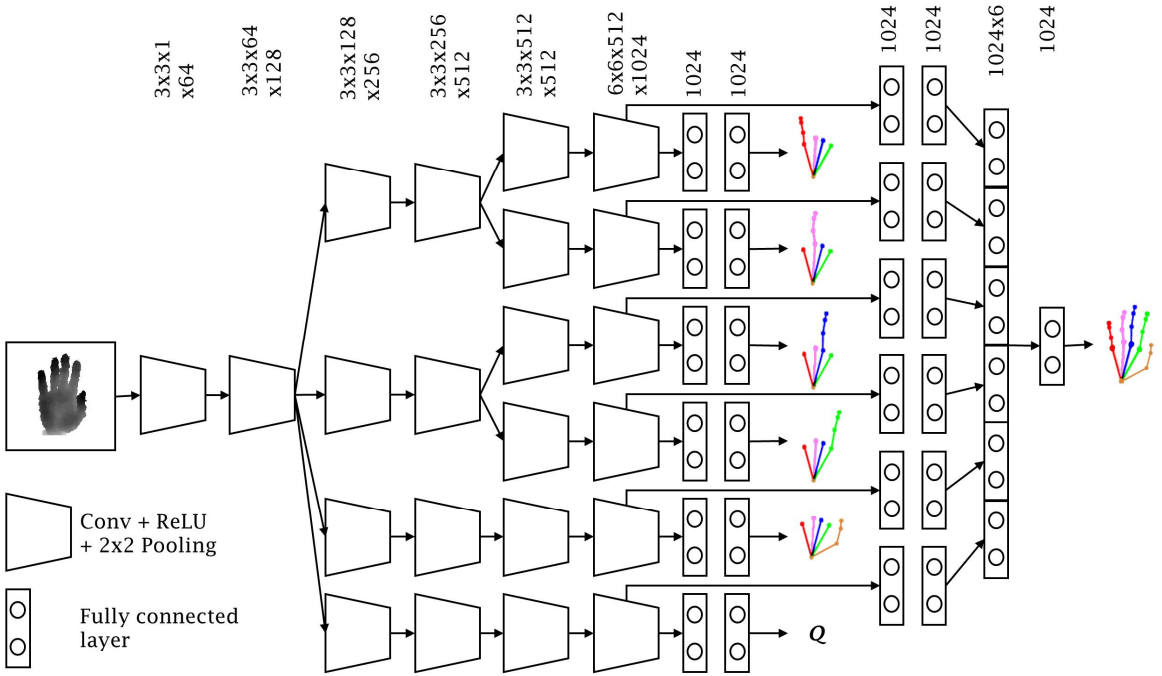
- CNNs can be used to learn joints heatmaps (Tompson 2014). However it is giving 2D pose,
- Multi-view fusion is an extension of heatmap-based methods for 3D pose (Ge 2016),
- Feature learning for direct pose regression in single channel network does not have enough capacity for complex poses and viewpoints,
- Regression over a linear embedded space of pose does not generalize well in practice (Oberweger CVWW2015),
- Generative error feedback models still generate model drifts (Oberweger ICCV2015),
- No specific constraints have been applied on the pose in the training process.

[1] Jonathan Tompson et al., Real-time continuous pose recovery of human hands using convolutional networks. TOG, 2014.
[2] Ge et al., Robust 3d hand pose estimation in single depth images: from single-view cnn to multi-view cnns. CVPR, 2016.
[3] Markus Oberweger et al., Hands deep in deep learning for hand pose estimation. Computer Vision Winter Workshop, 2015.
[4] Markus Oberweger et al., Training a feedback loop for hand pose estimation. ICCV, 2015.

CNN-based regressor

CNN architecture

- Hand pose is broken into a set of simpler sub-poses,
- Weights are shared in a hierarchy from general features to local features,
- Palm is modeled by a viewpoint regressor (*Q*),
- Local features are fused to generate global pose at the end.



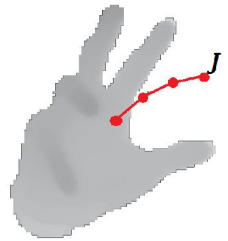
Constraints as loss function

- *L2* loss does not guarantee proper generalization of network,
- It is proved that *L2* loss is sensitive to the noise in the data,
- We accumulate *L2* loss with appearance and physical constraints,

$$L = \lambda_1 L_{loc} + \lambda_2 L_{glo} + \lambda_3 L_{app} + \lambda_4 L_{dyn}$$

- In appearance loss, all projected joints must have a larger depth than image pixels,

$$L_{app} = \sum_{i=1}^{m} \max(0, \mathscr{I}(j_i^u, j_i^v) - j_i^z)$$



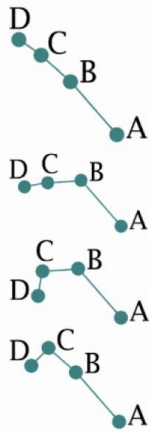
- Physical constraints are defined based on finger's dynamics:

case 1
$\|\overrightarrow{AB}\| + \|\overrightarrow{BC}\| + \|\overrightarrow{CD}\| < 1.01\|\overrightarrow{AD}\|$
$\overrightarrow{AB} \parallel \overrightarrow{AC} \parallel \overrightarrow{AD} \parallel \mathbf{e}_G$ ▷

case 2
$\overrightarrow{AB} \times \overrightarrow{BC} \parallel \overrightarrow{AC} \times \overrightarrow{CD} \parallel \mathbf{e}_G$ ▷

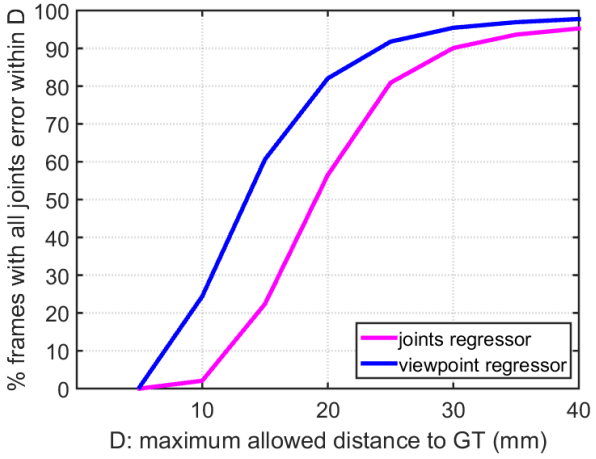
case 3
$\overrightarrow{AB} \times \overrightarrow{BC} \parallel \overrightarrow{BC} \times \overrightarrow{CD} \parallel \mathbf{e}_G$ ▷

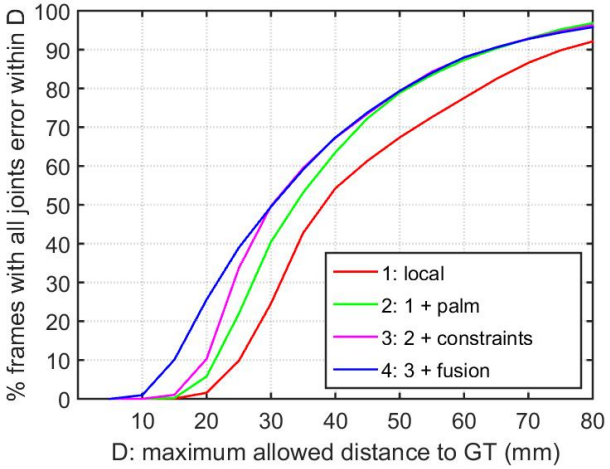
case 4
$\overrightarrow{AB} \times \overrightarrow{BC} \parallel \overrightarrow{AB} \times \overrightarrow{BD} \parallel \mathbf{e}_G$ ▷

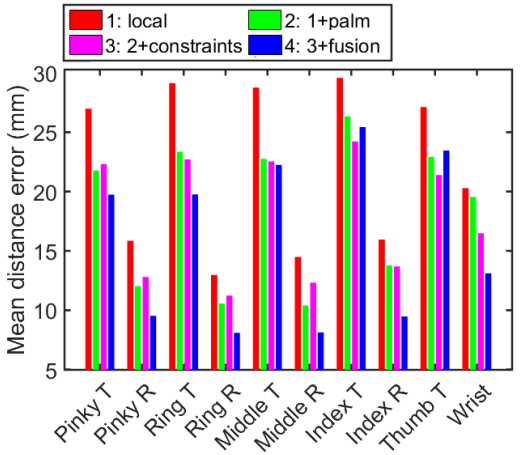


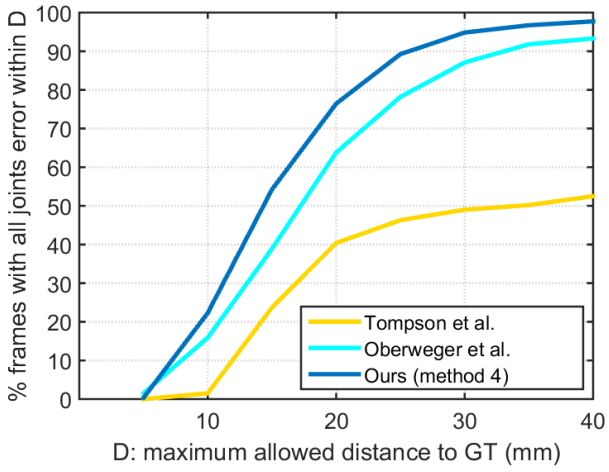
CNN-based regressor

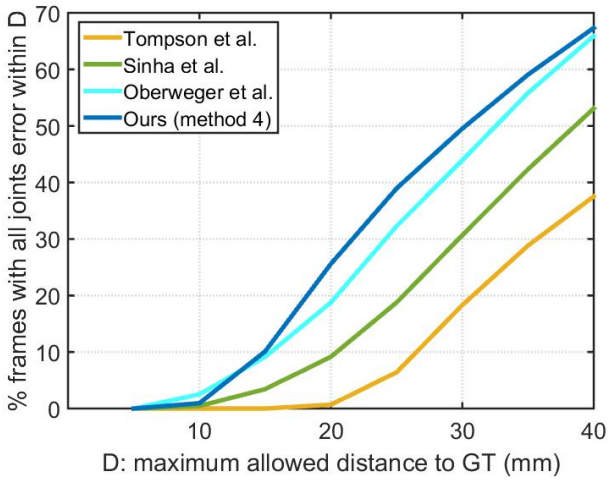
Results: NYU dataset











[1] Jonathan Tompson et al., Real-time continuous pose recovery of human hands using convolutional networks. TOG, 2014.
[2] Markus Oberweger et al., Training a feedback loop for hand pose estimation. ICCV, 2015.
[3] Ayan Sinha et al., Deephand: robust hand pose estimation by completing a matrix imputed with deep features. CVPR, 2016.

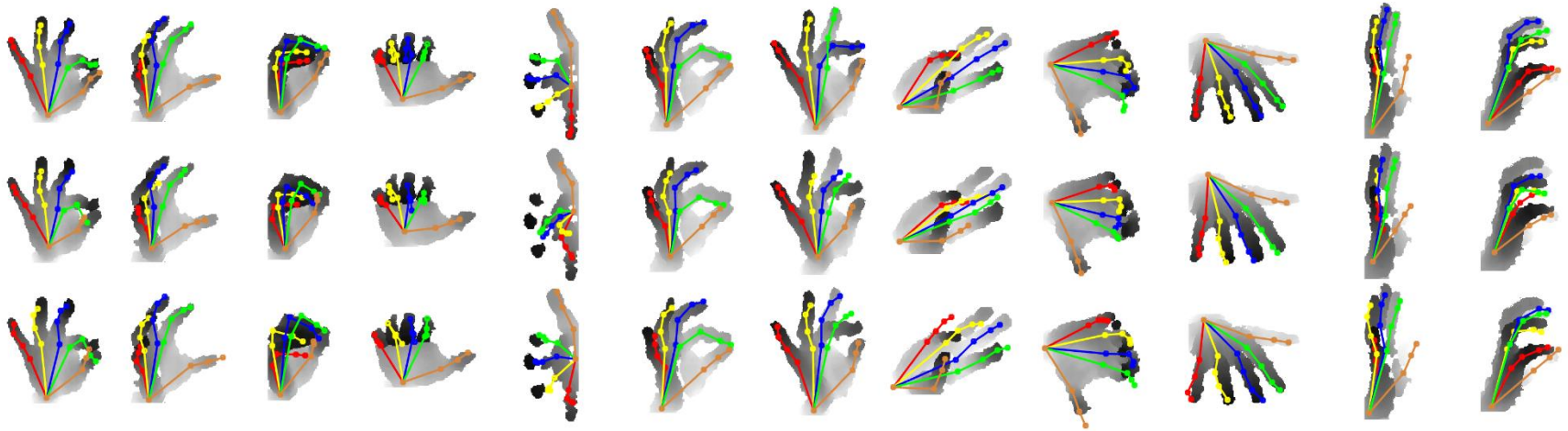
CNN-based regressor

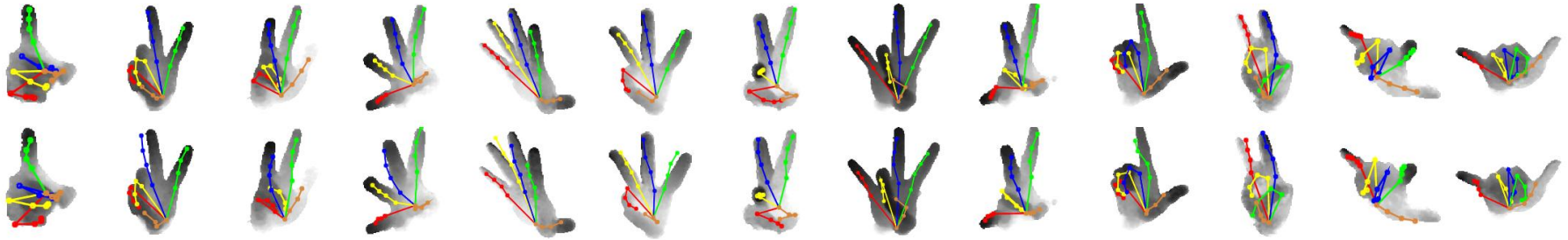
Results: NYU and MSRA datasets

Gt

Local

Global



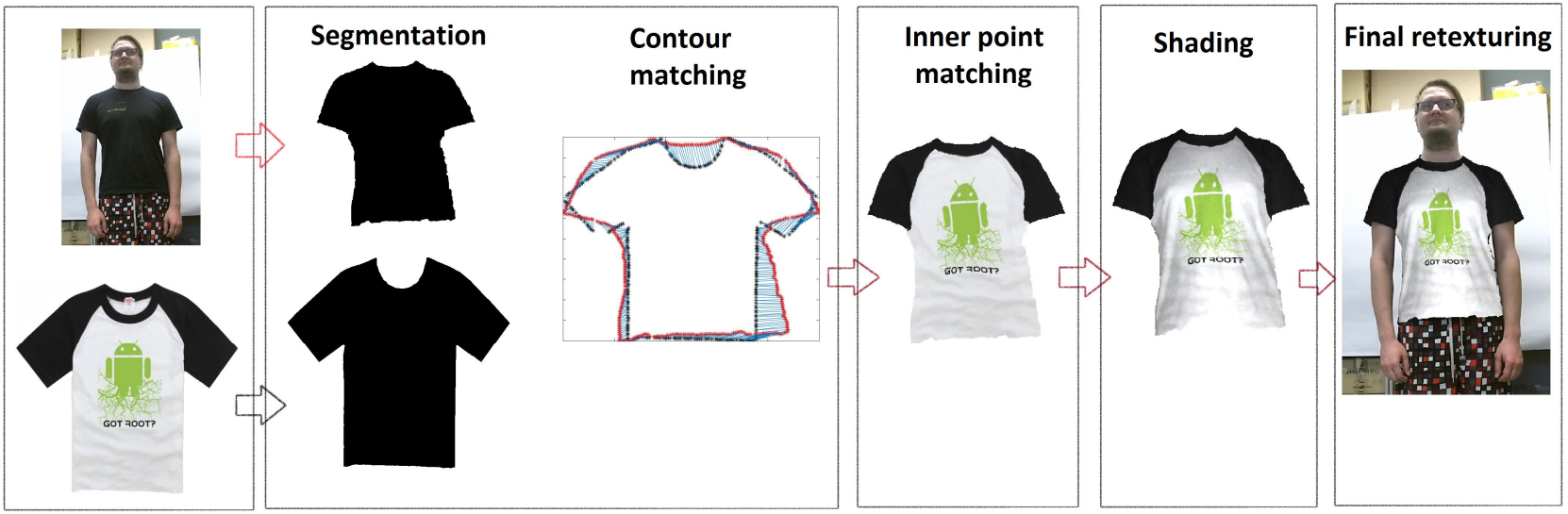


Introduction

- Garment retexturing is mainly used in retailing and/or movie editing,
- Problem is defined as mapping a 2D RGB texture to a 3D body surface, i.e. assigning each 3D point a color from flat garment.
- Challenges are included as:
  1. Possible occlusion of 3D surface,
  2. Inconsistency in the topology of surfaces,
  3. Shading new texture correctly.

System overview

- Retexturing method using RGB-D data is covered by
  1. Garment segmentation (by means of Grabcut),
  2. 2D to 3D garment matching (2D contour matching by GMM) and
  3. Rendering (by IR image for colors intensity).



[1] Carsten Rother et al., Grabcut: Interactive foreground extraction using iterated graph cuts. TOG, 2004.
[2] Bing Jian and Baba C. Vemuri. Robust point set registration using Gaussian mixture models. In PAMI, 2010.

2D to 3D garment matching

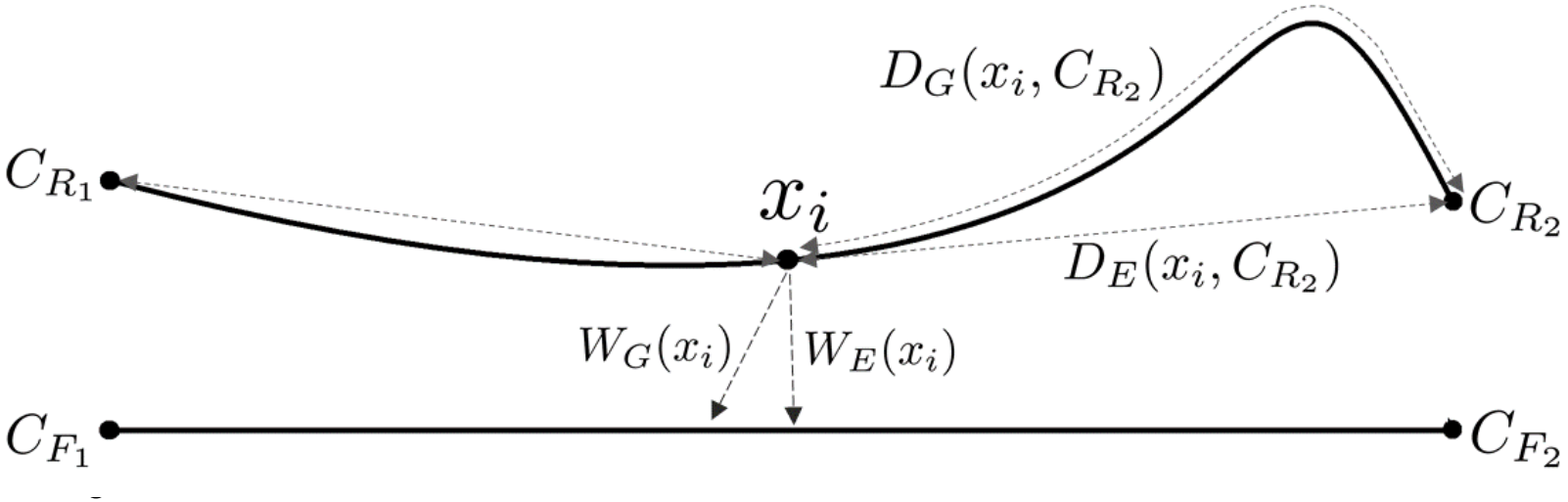
- We solve the problem by interpolating space between 2D and 3D garments.
- 2D garment deformation based on contours matching does not take surface topology into account,
- Thin plate spline can solve the problem in closed form.
- Given matched contours C_R and C_F, a mapping from 3D point x_i to RGB image is defined as:

$$W(x_i) = \sum_{j=1}^{n} \omega_j \kappa(\| x_i - C_{R_j} \|) \qquad \kappa(d) = d^2 \log d$$

- Radial basis kernel based on Euclidean distance does not take surface topology into account.
- Geodesic distance (fast marching algorithm) can solve the problem.

[1] Thomas Deschamps and Laurent D. Cohen. Fast extraction of minimal paths in 3d images and applications to virtual endoscopy, 2001

2D to 3D garment matching



$$W(x_i) = \sum_{j=1}^{n} \omega_j \kappa(\| x_i - C_{R_j} \|) \qquad \kappa(d) = d^2 \log d$$

- Radial basis kernel based on Euclidean distance does not take surface topology into account.
- Geodesic distance (fast marching algorithm) can solve the problem.

[1] Thomas Deschamps and Laurent D. Cohen. Fast extraction of minimal paths in 3d images and applications to virtual endoscopy, 2001

Dataset

- To evaluate our method, we created a dataset by Kinect2 consisting of
  - 91 RGB-D images of 14 individuals (11 males and 3 females) and 13 flat garments gathered from internet.
  - 39 RGB-D images of 5 individuals (4 males and 1 female) putting on 8 garments. Garments are attached 16 landmarks to evaluate real vs. retextured landmark locations.

Results

- To compute MOS, we showed 91 sets of images to 41 individuals to define the most realistic image among methods in comparison.

Mean Opinion Score (MOS) comparison

Method	T-shirt Votes	T-shirt Percentage	Long sleeve Votes	Long sleeve Percentage
NRICP	77	2.68%	32	3.69%
CPD	485	16.88%	245	28.23%
Ours	2311	80.44%	591	68.09%

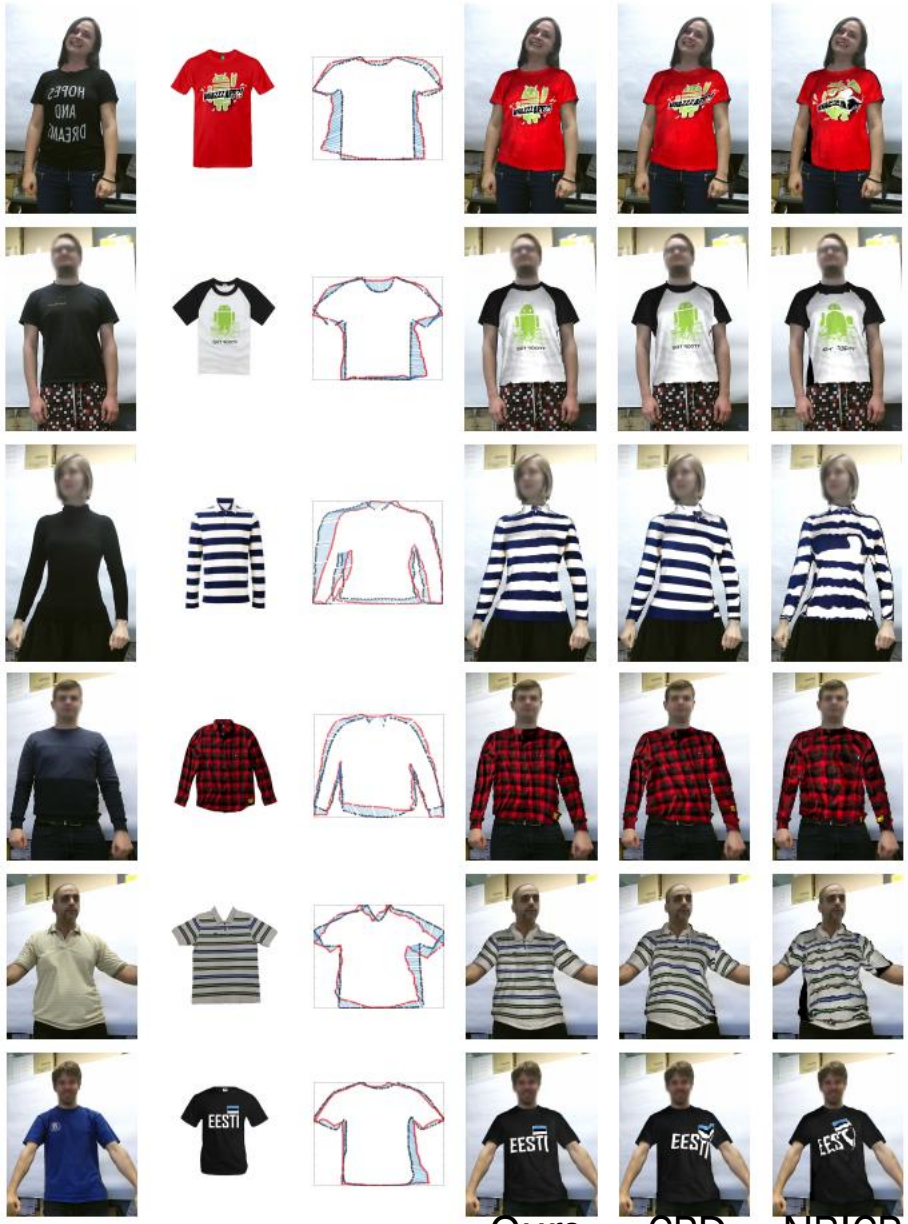
Marker mapping error

Method	MSE for T-shirts	MSE for Long sleeves
NRICP	115.400 px	215.349 px
CPD	83.850 px	190.618 px
Ours	75.005 px	105.884 px

[1] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid icp algorithms for surface registration. CVPR, 2007.
[2] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. PAMI, 2010.

Garment retexturing

Results



Body/hand segmentation

- We proposed nearest neighbor based solutions for human body/hand segmentation in depth images,
- We created a shape descriptor in depth images conditioning on each point class probability,
- We showed non-rigid model warping can generate accurate segmentation even for small segment regions.

- As future work:
  - TPS warping does not take mesh connections into account and can generate unrealistic shapes. Spring-like modeling may solve the problem.
  - Realistic and parametric models can be used as an alternative to avoid model drifts.

Face parsing

- We proposed an effective CNN architecture for face segmentation in RGB images in-the-wild,
- We modeled CRF as RNN able to learn pairwise kernels based on 4-connected graph
- We trained our CNN architecture end to end based on adversarial strategy,
- We showed conditioning the network on facial landmarks can improve results,
- We showed our model can accurately segment quite deformable face parts, e.g. lips and hair.

- We proposed a top-down generative strategy for hand pose recovery in depth images,
- We reduced search space by the aim of nearest neighbors,
- In a hierarchy palm is extracted and provide a basis for the finger model fitting,
- We incorporated spatio-temporal model for occlusion refinement,
- We showed our approach outperformed state of the art on complex datasets.

- As future work
  - In top-down strategies, error can be propagated from top to bottom. We will consider generative models jointly optimized with spatio-temporal models.

- We proposed a hierarchical CNN based solution for hand pose recovery in depth images,
- We trained local sub-poses jointly with global pose,
- We explicitly defined a loss by applying appearance and physical constraints on output joints,
- We showed a viewpoint regressor is more accurate than joint locations regressor for palm joints recovery,
- We showed our model outperformed state of the art on NYU and MSRA datasets.

- As future work
  - We will consider generative models and adversarial training for hand pose recovery.

- We developed an application for garment retexturing using RGB-D images in controlled situations,
- We solved 2D to 3D point matching by 2D contour matching as control points and 3D warping through TPS,
- We modeled surface topology by including geodesic distance in TPS,
- As a result, our model generated realistic retextured images on a gathered dataset.

- As future work
  - We will consider using parametric model fitting as an intermediate step for the applicability of garment retexturing in more complex body poses and occlusions.

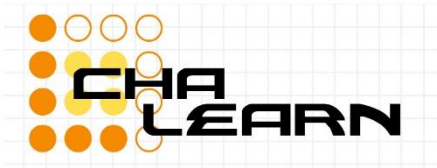
Publications

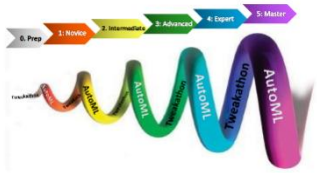
- Journal
  - Meysam Madadi, Egils Avots, Gholamreza Anbarjafari, Sergio Escalera, Xavier Baro, Jordi Gonzalez, From 2D to 3D Geodesic-based Garment Matching: A Virtual Fitting Room Approach, Under revision at IET Computer Vision, 2017.
  - Meysam Madadi, Sergio Escalera, Jordi Gonzàlez, F. Xavier Roca, Felipe Lumbreras, Multi-part body segmentation based on depth maps for soft biometry analysis, Pattern Recognition Letters 56 (2015), pp. 14–21
- Conference proceedings
  - Meysam Madadi, Sergio Escalera, Alex Carruesco Llorens, Carlos Andujar, Xavier Baro, Jordi Gonzalez, Occlusion aware hand pose recovery from sequences of depth images, 12th IEEE Conference on Automatic Face and Gesture Recognition (FG), 2017 A
  - Hugo Jair Escalante, Isabelle Guyon, Sergio Escalera, Julio Jacques Jr., Meysam Madadi, Xavier Baro, Stephane Ayache, Evelyne Viegas, Yagmur Gucluturk, Umut Guclu, Marcel van Gerven, Rob van Lier. Design of an Explainable Machine Learning Challenge for Video Interviews. Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN 2017), IEEE, 2017
  - Sergio Escalera, Xavier Baro, Jordi Gonzalez, Miguel A. Bautista, Meysam Madadi, Miguel Reyes, Víctor Ponce, Hugo J. Escalante, Jamie Shotton, Isabelle Guyon, ChaLearn Looking at People Challenge 2014: Dataset and Results, ChaLearn Looking at People, European Conference on Computer Vision, 2014. B
- ArXiv
  - Meysam Madadi, Sergio Escalera, Xavier Baro, Jordi Gonzalez, End-to-end Global to Local CNN Learning for Hand Pose Recovery in Depth data, arXiv:1705.09606, 2017.
  - Umut Guclu, Yagmur Gucluturk, Meysam Madadi, Sergio Escalera, Xavier Baro, Jordi Gonzalez, et al. End-to-end semantic face segmentation with conditional random fields as convolutional, recurrent and adversarial networks. Under revision at PAMI, 2017.

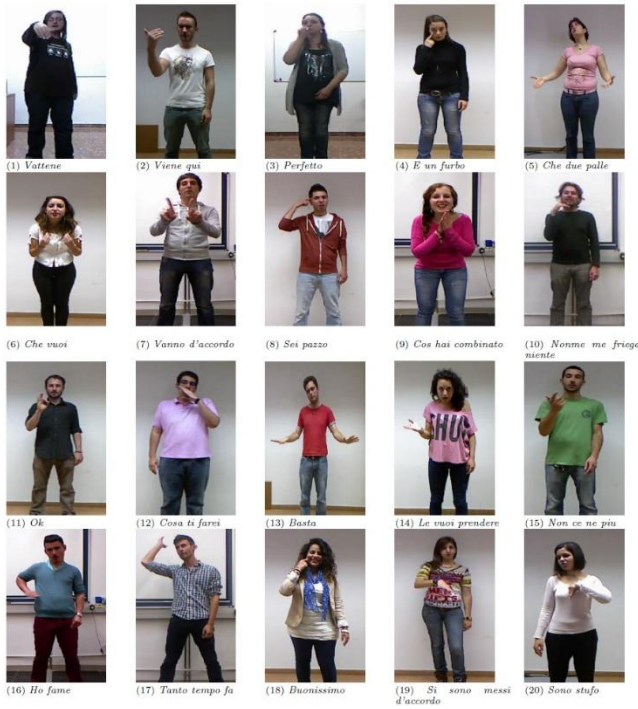
A 1 of the 18 selected papers for oral presentation among hundreds of submissions
B one of the current most used benchmarks for rgb-d gesture recognition

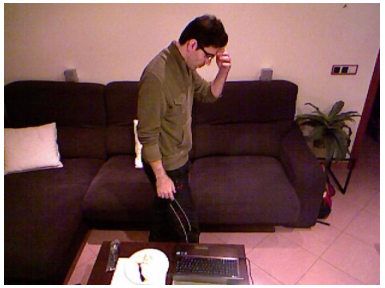
Projects

- Color correction in industrial printing,
- Elders monitoring,
- ChaLearn Looking at People,
- AutoML challenge,
- Fingerprint recognition demo,
- Ball detection in sport events.









Acknowledgement

- Finally, I would like to thank
  - My wife, Bahar, for all her patience and support,
  - My supervisors especially Dr. Sergio Escalera,
  - Generalitat de Catalunya for FIDGR fund,
  - ChaLearnLAP and Prof. Isabelle Guyon,
  - And all of you for your attention.

This work has been partially supported by the Spanish project TIN2016-74946-P (MINECO/FEDER, UE) and CERCA Programme / Generalitat de Catalunya. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.