UNDERSTANDING HUMAN-CENTRIC IMAGES

From Geometry to Fashion

Edgar Simo-Serra

Barcelona, 6th of July, 2015

Institut de Robòtica i Informàtica Industrial
CSIC UPC

MOTIVATION

MOTIVATION

MOTIVATION


Kimono
着物
Shiromuku
白無垢
Hakama
袴

MOTIVATION

Institut de Robòtica i Informàtica Industrial

MOTIVATION

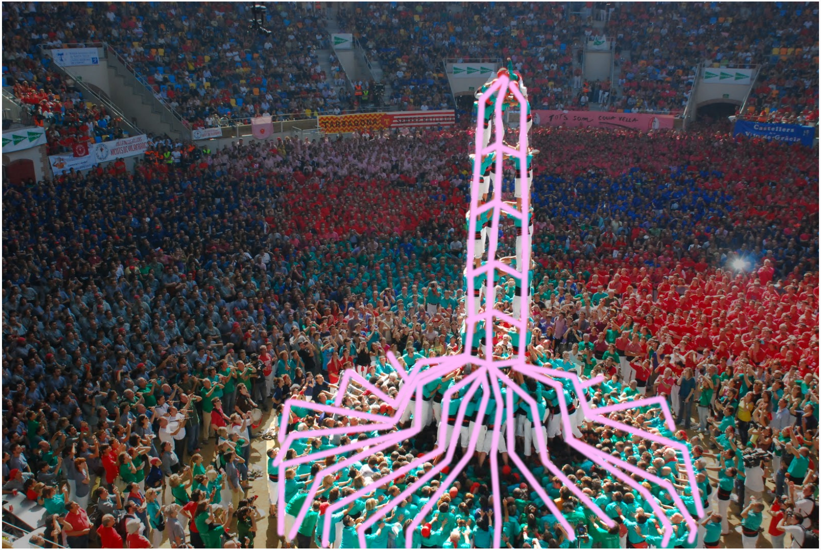
MOTIVATION

MOTIVATION

MOTIVATION

Institut de Robòtica i Informàtica Industrial

MOTIVATION



MOTIVATION

MOTIVATION

MOTIVATION

MOTIVATION


- Lots of information can be obtained from a single image
  - Location, context, roles, relationships, . . .
- Prior knowledge is necessary
- Must build frameworks from the ground up
- This thesis is an effort towards higher level image understanding

Overview

Feature Point Descriptors

- Deformation and Light Invariant (DaLI) Descriptor
- Deep Convolutional Neural Network Descriptors

Generative 3D Human Pose Models

- Linear Latent Models
- Directed Acyclic Graphs
- Geodesic Finite Mixture Models

3D Human Pose Estimation

- 3D Pose Estimation from Noisy Observations
- Joint 2D and 3D Pose Estimation

Fashion Understanding

- Semantic Segmentation of Clothing
- Modelling Fashionability

Conclusions

OVERVIEW

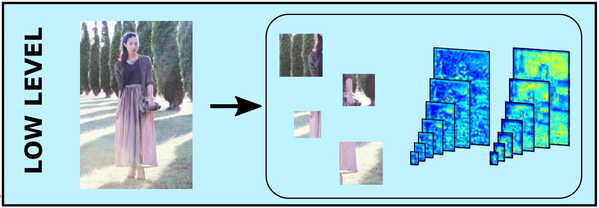
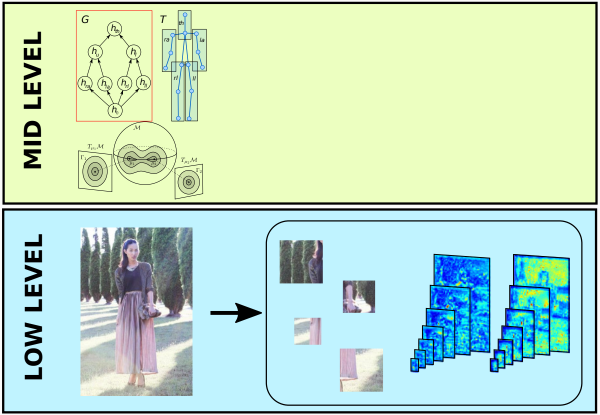


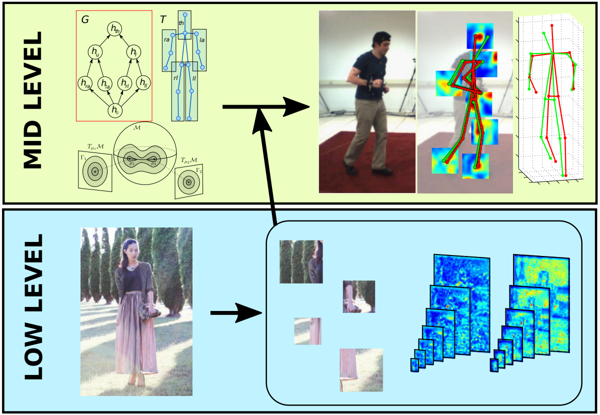
Image Description

OVERVIEW



Prior Models

Image Description

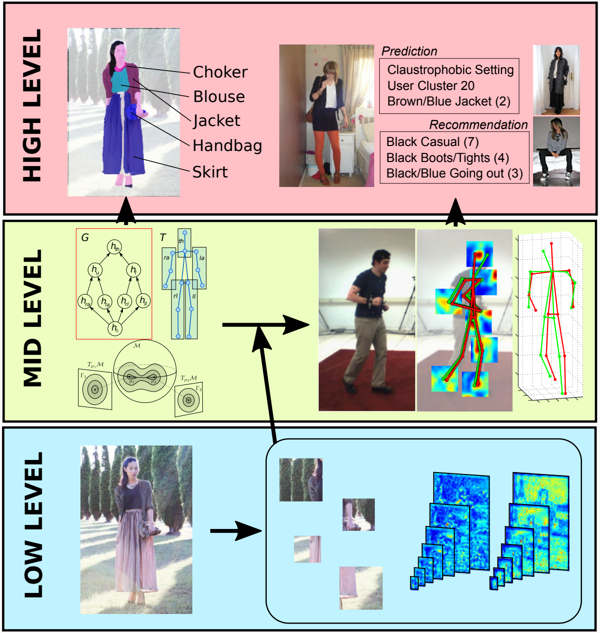


Pose Estimation

Prior Models

Image Description

OVERVIEW



Fashion Understanding

Semantic Segmentation

Pose Estimation

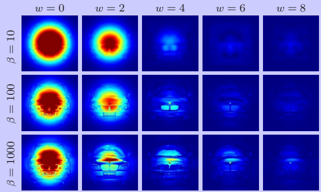
Prior Models

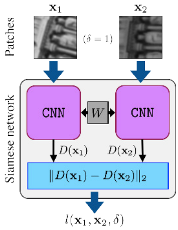
Image Description

FEATURE POINT DESCRIPTORS

DEFORMATION AND LIGHT INVARIANT (DALI) DESCRIPTOR

Deformation and Light Invariant (DaLI) Descriptor





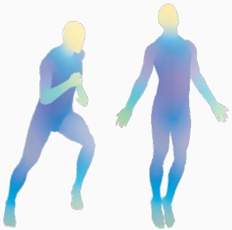
Deep Convolutional Neural Network Descriptors

Problem: matching points of interest under:

- Non-rigid deformations
- Photometric changes

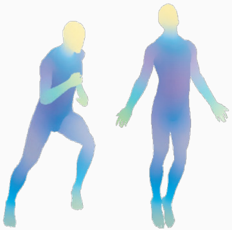
Diffusion of heat for 3D mesh matching

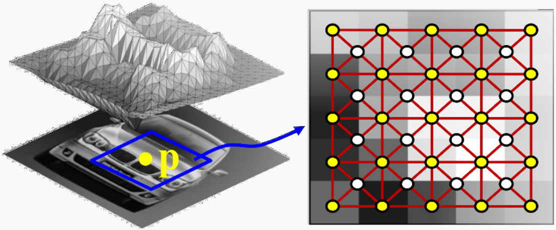
- Invariant to isometries
- Solution is given by **Heat Kernel Signature** [1]
- Time interval corresponds to globalness of the description



[1] J. Sun, M. Ovsjanikov, L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In Eurographics Symposium on Geometry Processing, 2009.

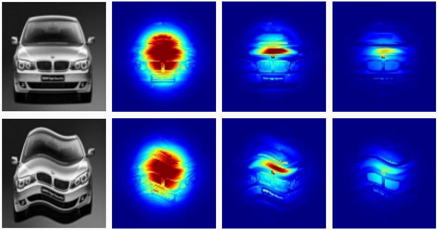
- Diffusion of heat for 3D mesh matching
  - Invariant to isometries
  - Solution is given by **Heat Kernel Signature** [1]
  - Time interval corresponds to globalness of the description
- Embed the image as a 3D surface to apply to images





[1] J. Sun, M. Ovsjanikov, L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In Eurographics Symposium on Geometry Processing, 2009.

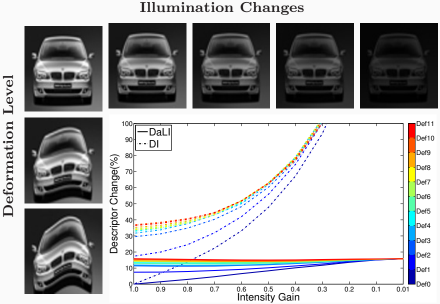
- Diffusion of heat for 3D mesh matching
  - Invariant to isometries
  - Solution is given by **Heat Kernel Signature** [1]
  - Time interval corresponds to globalness of the description
- Embed the image as a 3D surface to apply to images
- Heat diffusion along the surface is used as a descriptor



[1] J. Sun, M. Ovsjanikov, L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In Eurographics Symposium on Geometry Processing, 2009.

HKS is sensitive to scale (illumination)

Use Fast Fourier Transform to gain invariance to scale [2]



[2] M. Bronstein, I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In CVPR, 2010.

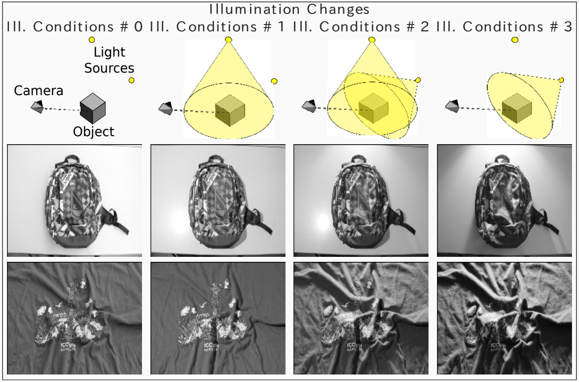
DALI DATASET

- New deformation and illumination dataset
- 12 objects, 4 deformation levels, 4 illumination levels
- Manual annotation of correspondences

New deformation and illumination dataset

12 objects, 4 deformation levels, 4 illumination levels

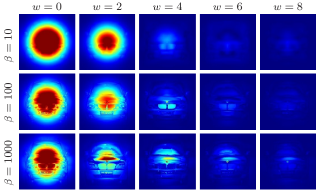
Manual annotation of correspondences

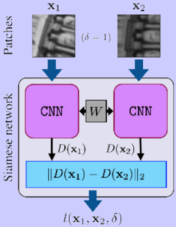


DaLI outperforms all, especially in illumination

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

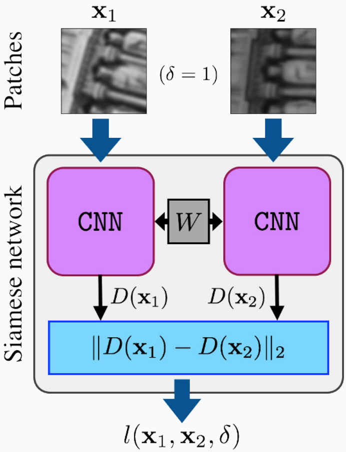
Deformation and Light Invariant (DaLI) Descriptor





Deep Convolutional Neural Network Descriptors

- Learn using pairs of patches jointly
- Minimize distance for ”same” patches, maximize for ”different” patches
- Need to use all the tricks to get good performance



Institut de Robòtica i Informàtica Industrial

Implemented in Torch7 (Lua, LeCun et al.)

INPUT: 64x64 patch

3 Convolutional layers, 46,272 parameters

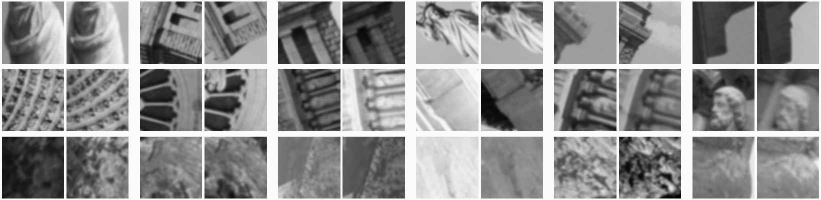
1. 32x7x7 Kernel, Tanh, 2x pooling, normalization
2. 64x6x6 Kernel, Tanh, 3x pooling, normalization
3. 128x5x5 Kernel, Tanh, 4x pooling

OUTPUT: 128 dimension vector

DATA

Learn on Structure from Motion dataset

Ground truth created by using 3D structure

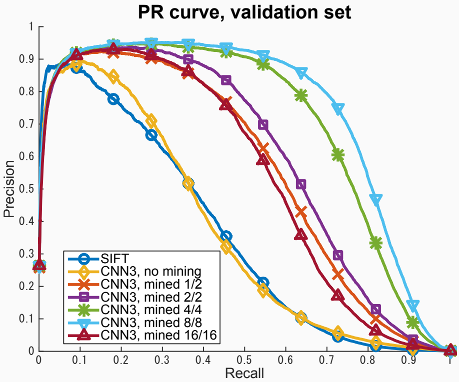


Sampling approach for negatives/positives

- 10^6 positives, 10^{12} negatives

Large amounts of mining

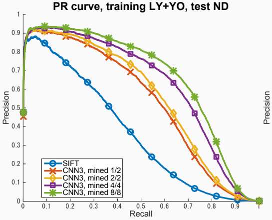
Essential for performance

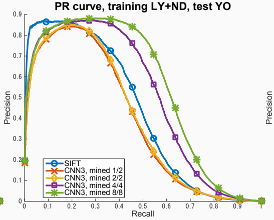


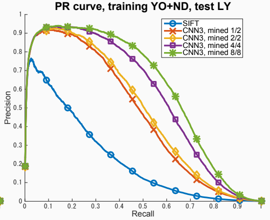
Institut de Robòtica i Informàtica Industrial

RESULTS

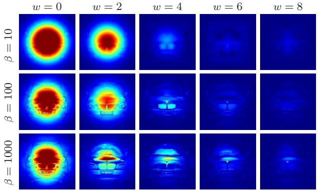
Train	Test	SIFT	CNN3	PR AUC Increase
LY+YOS	ND	0.349	**0.667**	91.1%
LY+ND	YOS	0.425	**0.545**	28.2%
YOS+ND	LY	0.226	**0.608**	169.0%

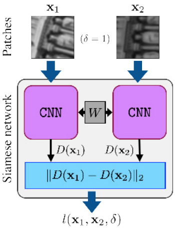






Deformation and Light Invariant (DaLI) Descriptor

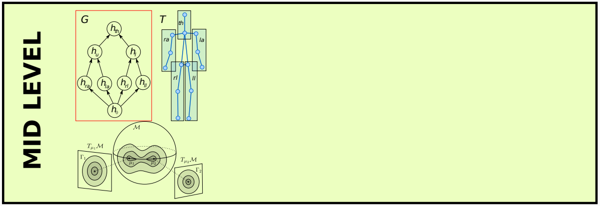




Deep Convolutional Neural Network Descriptors

- SIFT is ubiquitous in computer vision
  - Better alternatives out there (DAISY, DaLI, . . .)
- Alternative descriptions can be complementary
- Trend to move away from hand-crafted features to learnt features continues

3D human pose models



Prior Models

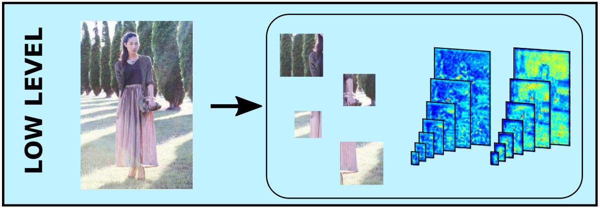
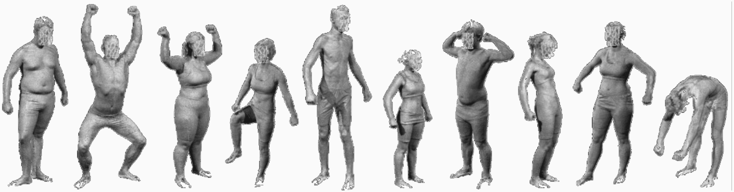


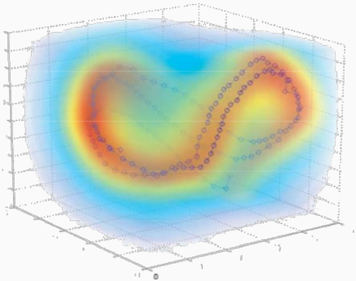
Image Description

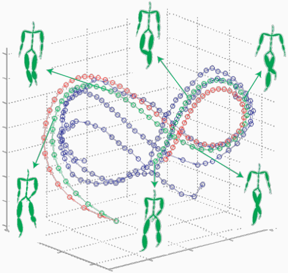
OVERVIEW

Shape Models (Hasler et al., 2009)



GPDM Models (Wang et al., 2005)



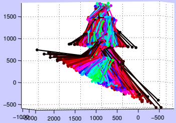


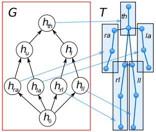
Institut de Robòtica i Informàtica Industrial

Overview of different generative 3D human pose models

Model	Complexity	Scales?	Consistent?	PDF?
GMM	Low	Yes	No	Yes
PGA	Low	Yes	Yes	No
GPLVM	Low	No	No	Yes
GPDM	Medium	No	No	Yes
hGPLVM	Medium	No	No	Yes
CRBM	High	Yes	No	Yes
GCMFA	High	No	No	Yes
PCA	Low	Yes	No	No
DAG	Medium	Yes	No	Yes
GFMM	Low	Yes	Yes	Yes

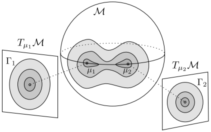
Linear Latent Model (PCA)





Directed Acyclic Graphs (DAG)

Geodesic Finite Mixture Models (GFMM)

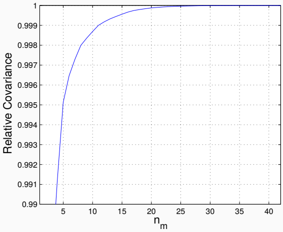


Represent pose as a linear combination of deformation bases

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

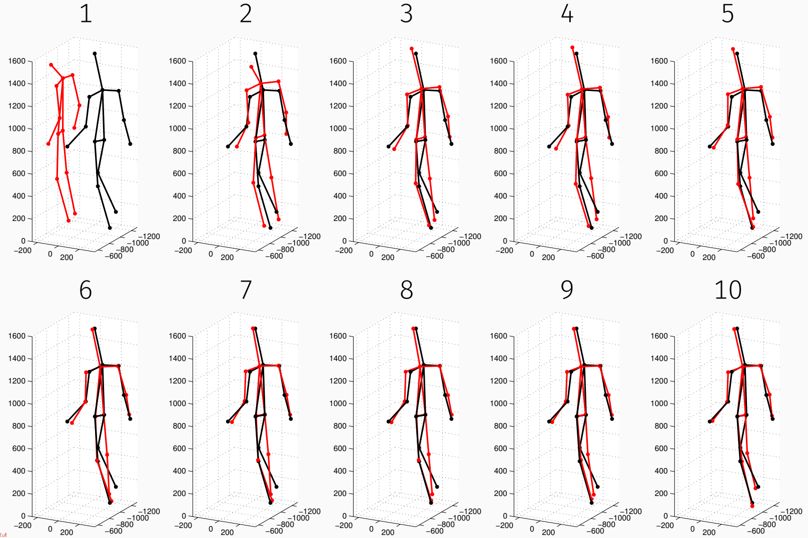
Bases found by computing SVD on the covariance of training data

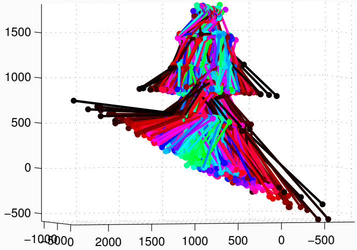
n_m eigenvectors corresponding to largest eigenvalues as basis



LINEAR LATENT MODELS (PCA)

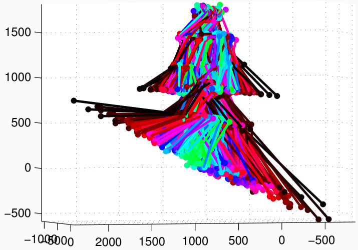
Dimension of the Latent Space n_m





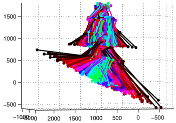
Very fast to both train and use

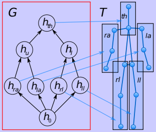
Linear formulation



- Very fast to both train and use
- Linear formulation
- Not probabilistic
  - Can generate non-anthropomorphic poses

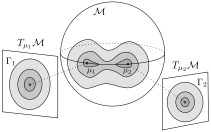
Linear Latent Model (PCA)





Directed Acyclic Graphs (DAG)

Geodesic Finite Mixture Models (GFMM)

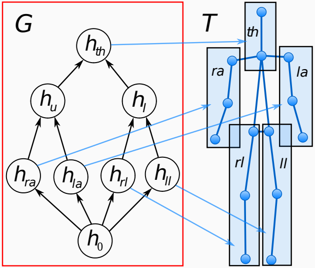


Institut de Robòtica i Informàtica Industrial

Model pose with a graphical model

Probabilistically encode plausible configurations

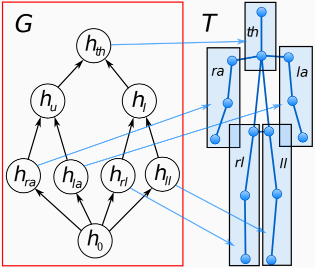
Directed Acyclic Graph allows for dynamic programming



Model pose with a graphical model

Probabilistically encode plausible configurations

Directed Acyclic Graph allows for dynamic programming



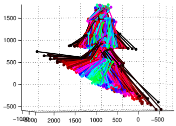
Poses discreticized with k-means clustering

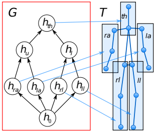
Discrete locations associated with latent states

Learnt using maximum likelihood

Efficient functions that map from latent space to pose and back

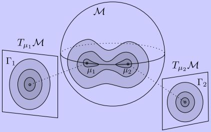
Linear Latent Model (PCA)





Directed Acyclic Graphs (DAG)

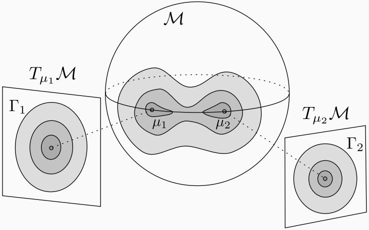
Geodesic Finite Mixture Models (GFMM)



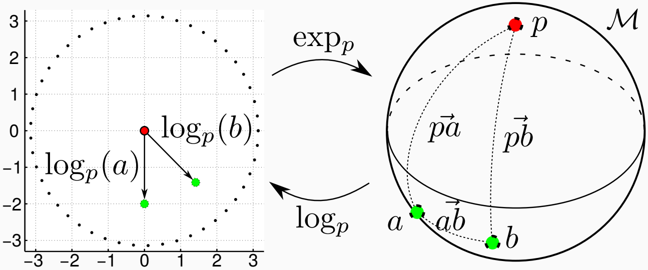
Institut de Robòtica i Informàtica Industrial

OVERVIEW

- Model Probability Density Function (PDF) of data on a manifold
- Fully unsupervised algorithm
- Efficient implementation
- One tangent space per cluster



Geodesic distance between two points on a manifold is the shortest distance along the manifold

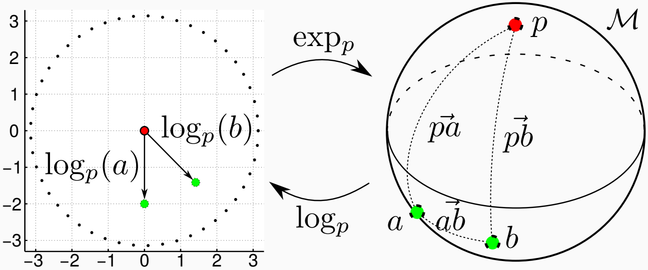


Institut de Robòtica i Informàtica Industrial

Geodesic distance between two points on a manifold is the shortest distance along the manifold

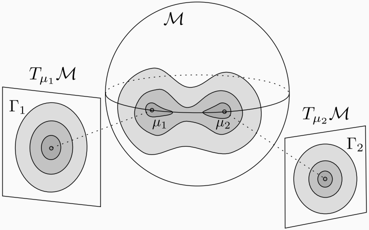
Tangent space is a local approximation of a manifold that is a **Euclidean space**

- logarithm and exponential map project to and from a tangent space respectively



Expectation-Maximization algorithm

- **Minimum Message Length** used to determine number of clusters
- Random initialization with large amount of clusters

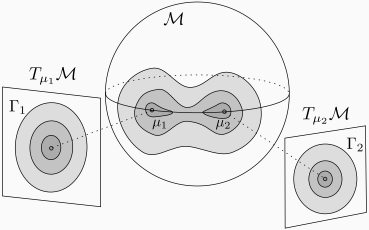


Expectation-Maximization algorithm

- **Minimum Message Length** used to determine number of clusters
- Random initialization with large amount of clusters

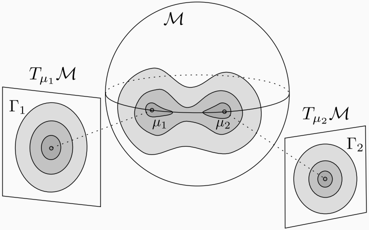
Expectation

- Data softly assigned to clusters

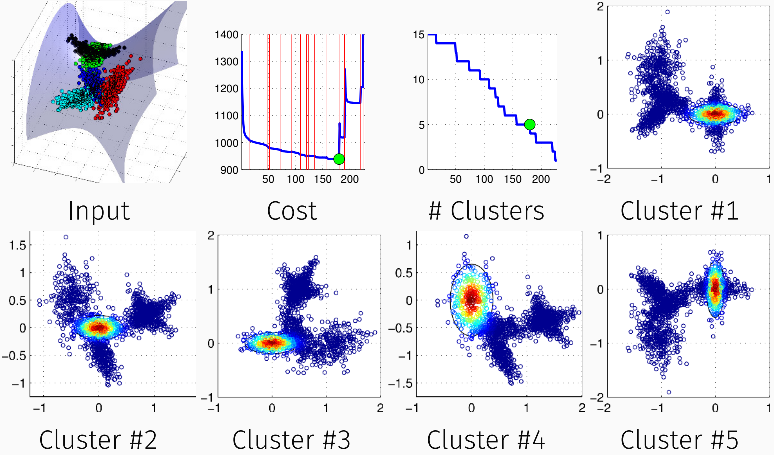


Institut de Robòtica i Informàtica Industrial

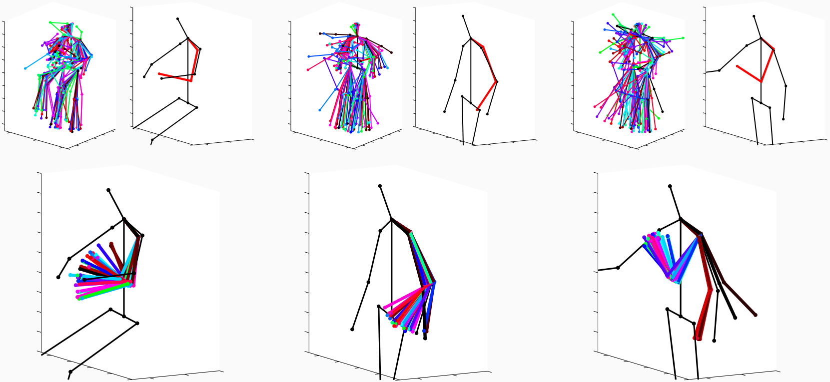
- Expectation-Maximization algorithm
  - **Minimum Message Length** used to determine number of clusters
  - Random initialization with large amount of clusters
- Expectation
  - Data softly assigned to clusters
- Maximize probability with tangent spaces
  - Mean estimated on the manifold using the **geodesic mean**
  - Covariance estimated on the **tangent space** in closed form



RESULTS - SYNTHETIC

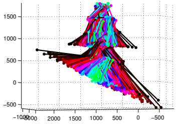


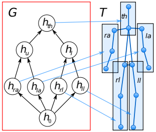
Regression gives another GFMM



Scenario 1 Scenario 2 Scenario 3

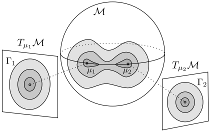
Linear Latent Model (PCA)





Directed Acyclic Graphs (DAG)

Geodesic Finite Mixture Models (GFMM)



Overview of different generative 3D human pose models

Model	Complexity	Scales?	Consistent?	PDF?
PCA	Low	Yes	No	No
DAG	Medium	Yes	No	Yes
GFMM	Low	Yes	Yes	Yes

Many different ways to model the pose

Each models has different strengths/weaknesses

Exploiting known properties is beneficial

Overview of different generative 3D human pose models

Model	Complexity	Scales?	Consistent?	PDF?
PCA	Low	Yes	No	No
DAG	Medium	Yes	No	Yes
GFMM	Low	Yes	Yes	Yes

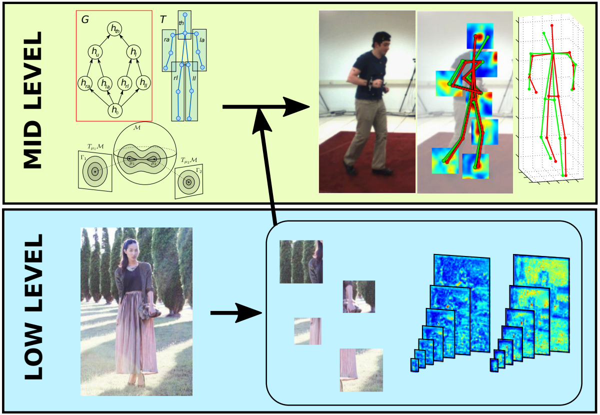
Many different ways to model the pose

Each models has different strengths/weaknesses

Exploiting known properties is beneficial

Now to models in action!

3D human pose estimation



Pose Estimation

Prior Models

Image Description

GIVEN:

- Single input image
- Internal calibration A

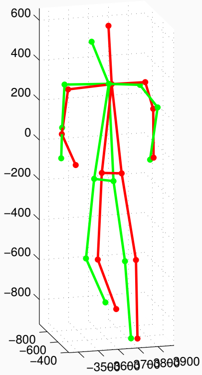
GIVEN:

Single input image

Internal calibration A

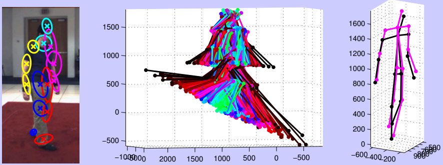
OBJECTIVE:

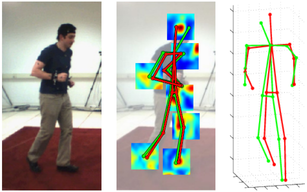
Retrieve 3D pose



3D POSE ESTIMATION FROM NOISY OBSERVATIONS

Single Image 3D Human Pose Estimation from Noisy Observations

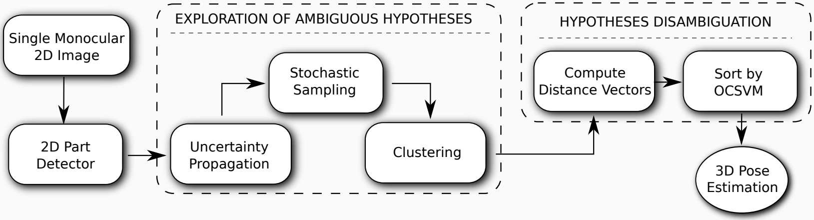




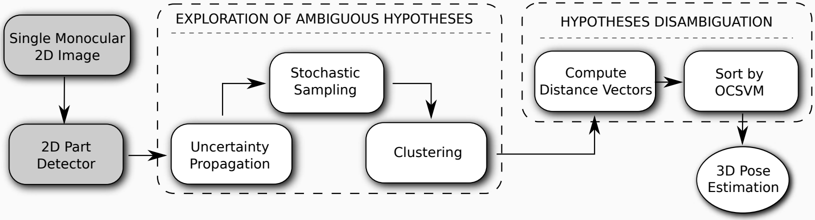
A Joint Model for 2D and 3D Pose Estimation from a Single Image

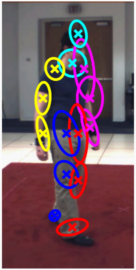
Institut de Robòtica i Informàtica Industrial

3D POSE ESTIMATION FROM NOISY OBSERVATIONS

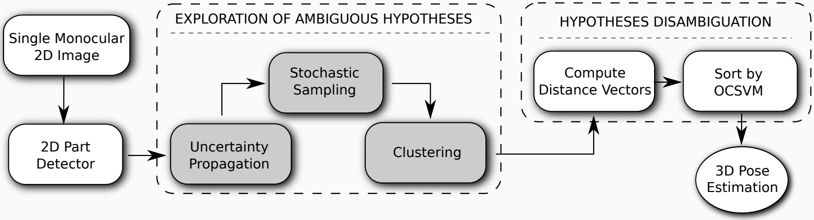


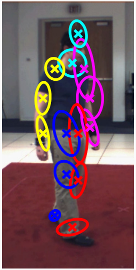
3D POSE ESTIMATION FROM NOISY OBSERVATIONS

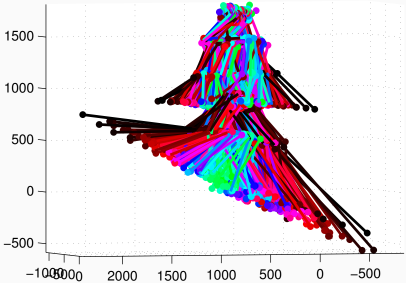




3D POSE ESTIMATION FROM NOISY OBSERVATIONS

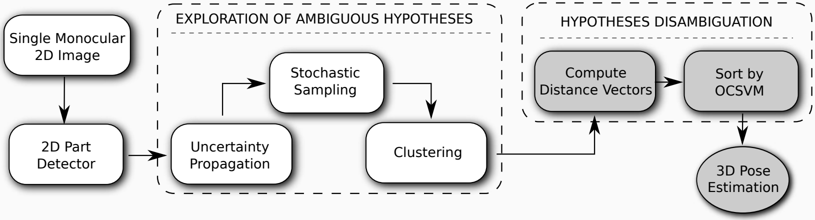


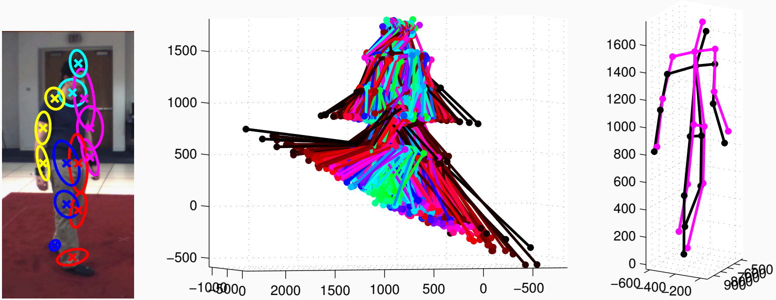




Institut de Robòtica i Informàtica Industrial

3D POSE ESTIMATION FROM NOISY OBSERVATIONS





Projective Linear Deformation Model

Camera projection can be written as a linear equation

$$Mx = 0$$

Projective Linear Deformation Model

- Camera projection can be written as a linear equation
- Principal component analysis is also a linear equation

$$\begin{aligned} Mx &= 0 \\ x &= x_0 + Q\alpha \end{aligned}$$

Institut de Robòtica i Informàtica Industrial

Projective Linear Deformation Model

- Camera projection can be written as a linear equation
- Principal component analysis is also a linear equation
- Rank deficient system

$$\left.\begin{aligned} Mx &= 0 \\ x &= x_0 + Q\boldsymbol{\alpha} \end{aligned}\right\} \implies MQ\boldsymbol{\alpha} + Mx_0 = 0$$

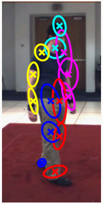
UNCERTAINTY PROPAGATION

Projective Linear Deformation Model

- Camera projection can be written as a linear equation
- Principal component analysis is also a linear equation
- Rank deficient system

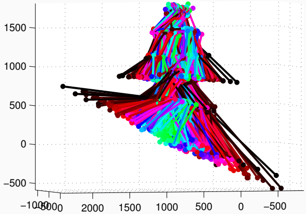
$$\left.\begin{aligned} Mx &= 0 \\ x &= x_0 + Q\boldsymbol{\alpha} \end{aligned}\right\} \implies MQ\boldsymbol{\alpha} + Mx_0 = 0$$

2D Gaussians propagated through linear system to poses

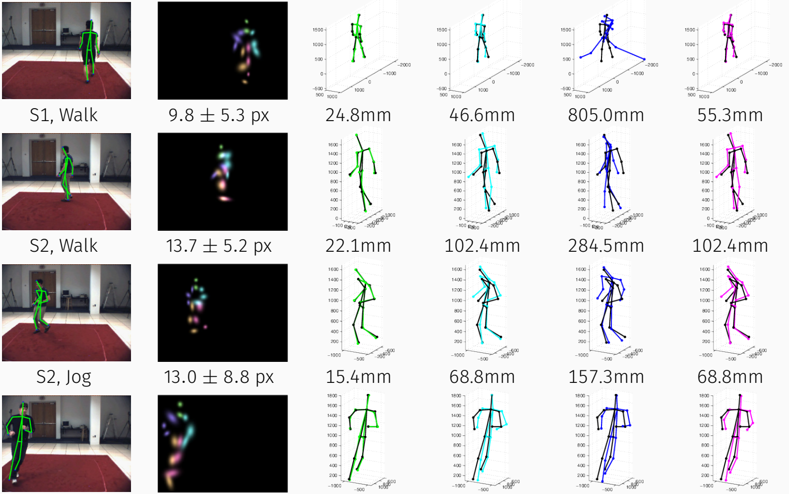


$$\boldsymbol{\mu}_\alpha = -(MQ)^\dagger Mx_0$$

$$\Sigma_\alpha = \frac{\delta\alpha}{\delta\boldsymbol{\mu}}\Sigma_\mu\left(\frac{\delta\alpha}{\delta\boldsymbol{\mu}}\right)^\mathsf{T}$$



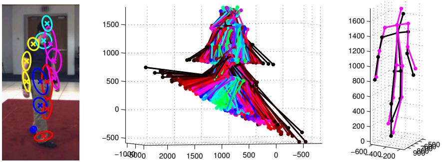
RESULTS

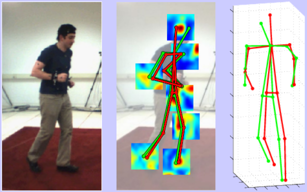


G. Truth	Detection	Opt. PCA	Best Rec.	Best Err.	Ours
S1, Walk	9.8 ± 5.3 px	24.8mm	46.6mm	805.0mm	55.3mm
S2, Walk	13.7 ± 5.2 px	22.1mm	102.4mm	284.5mm	102.4mm
S2, Jog	13.0 ± 8.8 px	15.4mm	68.8mm	157.3mm	68.8mm
S3, Jog	17.5 ± 13.0 px	26.3mm	72.6mm	105.3mm	89.5mm

Institut de Robòtica i Informàtica Industrial

Single Image 3D Human Pose Estimation from Noisy Observations

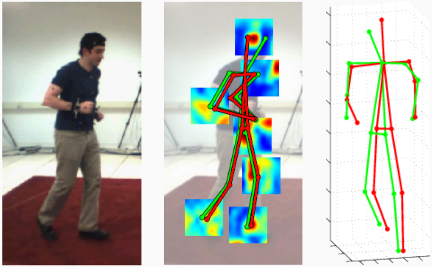




A Joint Model for 2D and 3D Pose Estimation from a Single Image

Propose single framework for 2D and 3D

Probabilistic extendible framework



Consider image evidence to be independent for each part:

$$p(D \mid L) = \prod_{i=1}^{N} p(\boldsymbol{d}_i \mid \boldsymbol{l}_i)$$

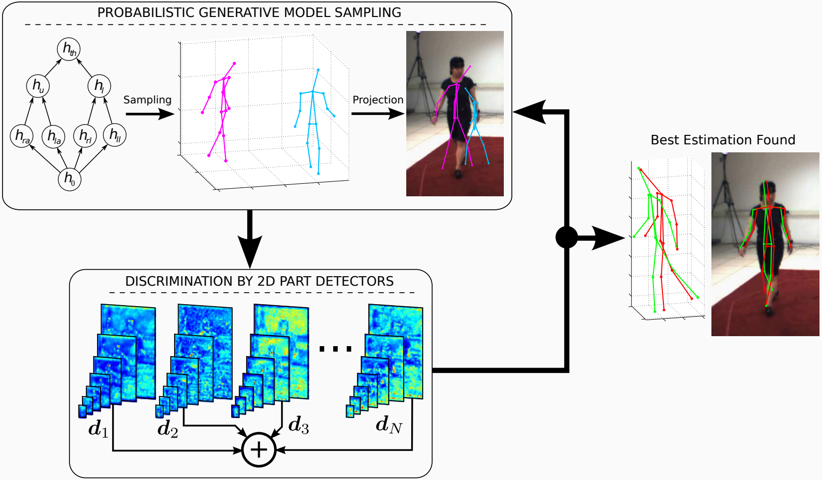
Consider image evidence to be independent for each part:

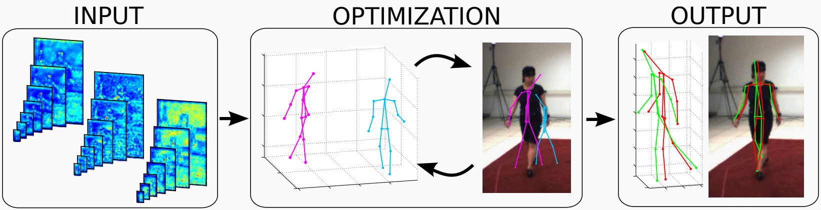
$$p(D \mid L) = \prod_{i=1}^{N} p(\boldsymbol{d}_i \mid \boldsymbol{l}_i)$$

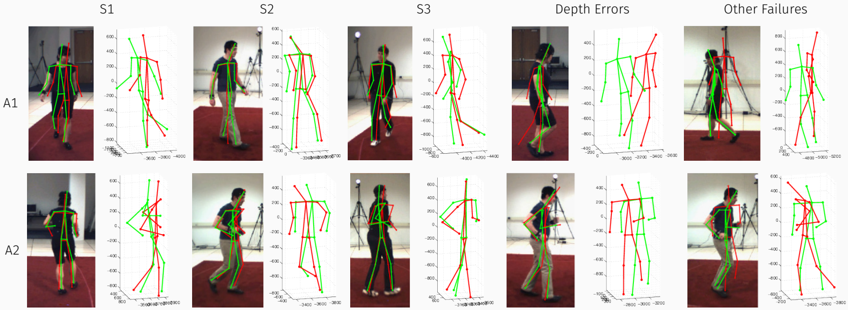
Bayes' rule and consider $p(L) = p(L \mid X)\, p(X \mid H)\, p(H)$

$$p(X \mid D) \propto \underbrace{p(H)\, p(X \mid H)}_{\text{generative}} \underbrace{\prod_{i=1}^{N} \left(p(\boldsymbol{d}_i \mid \boldsymbol{l}_i)\, p(\boldsymbol{l}_i \mid \boldsymbol{x}_i)\right)}_{\text{discriminative}}$$

BAYESIAN FORMULATION

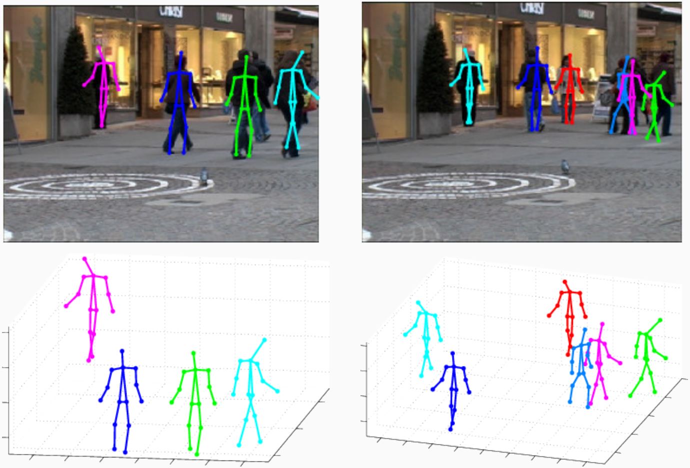






Institut de Robòtica i Informàtica Industrial

RESULTS - TUD STADMITTE



	Walking (A1,C1)		
	S1	S2	S3
Joint Model	65.1 (17.4)	48.6 (29.0)	73.5 (21.4)
Noisy Observations	99.6 (42.6)	108.3 (42.3)	127.4 (24.0)
[1] (tracking)	89.3	108.7	113.5
[2] (tracking)	-	107 (15)	-
[3] (background subtraction)	38.2 (21.4)	32.8 (23.1)	40.2 (23.2)
	Jogging (A2,C1)		
	S1	S2	S3
Joint Model	74.2 (22.3)	46.6 (24.7)	32.2 (17.5)
Noisy Observations	109.2 (41.5)	93.1 (41.1)	115.8 (40.6)
[3] (background subtraction)	42.0 (12.9)	34.7 (16.6)	46.4 (28.9)

[1] M. Andriluka, S. Roth, B. Schiele. Monocular 3d pose estimation and tracking by detection. In CVPR, 2010.

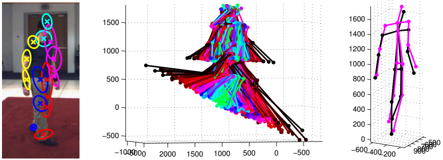
[2] B. Daubney, X. Xie. Tracking 3d human pose with large root node uncertainty. In CVPR, 2011.

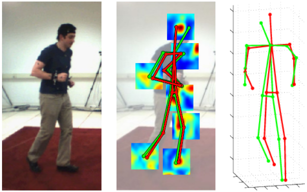
[3] L. Bo, C. Sminchisescu. Twin Gaussian Processes for Structured Prediction. IJCV, 87(1-2): 28-52, 2010.

Institut de Robòtica i Informàtica Industrial

3D POSE ESTIMATION

Single Image 3D Human Pose Estimation from Noisy Observations



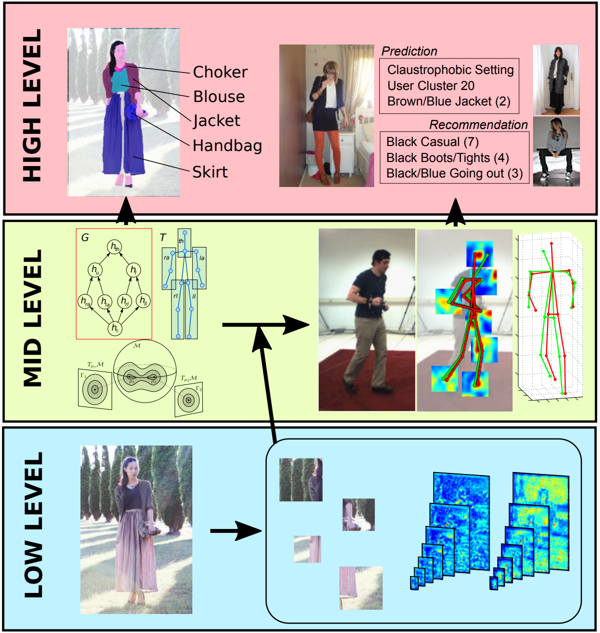


A Joint Model for 2D and 3D Pose Estimation from a Single Image

- Single image 3D pose estimation is an ambiguous problem
- 2D evidence is very unreliable
- Strong models necessary for performance
- Joint models perform best
    - Can exploit information
    - Delay decision until the end

FASHION UNDERSTANDING

OVERVIEW



Fashion Understanding

Semantic Segmentation

Pose Estimation

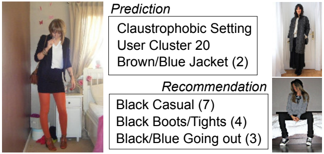
Prior Models

Image Description

Institut de Robòtica i Informàtica Industrial

SEMANTIC SEGMENTATION OF CLOTHING

Semantic Segmentation of Clothing



Modelling Fashionability

Semantic segmentation of clothing garments

Large inter and intra class variability

Fine-grained recognition task

background	heels	blazer	stockings	blouse	hat
hair	wedges	cardigan	tights	top	purse
skin	shoes	jumper	shorts	skirt	belt

Institut de Robòtica i Informàtica Industrial

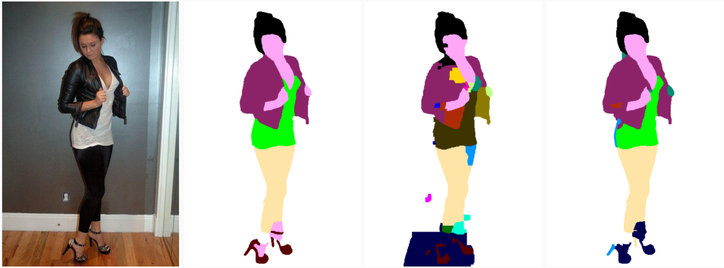
CONTRIBUTIONS

30% over state-of-the-art performance

Novel potentials that exploit the task

Efficient model that dresses the person

Input Truth [1] Ours



[1] K. Yamaguchi, H. Kiapour, L.E. Ortiz, T.L. Berg. Parsing clothing in fashion photographs. In CVPR, 2012.

- Superpixels labels
  $y_i \in \{1, \ldots, C\}$
- Limb segment labels
  $l_p \in \{1, \ldots, C\}$
- Unary potentials
  - Simple features
  - Person Mask
  - Clothelets
  - Shape features
- Pairwise potentials
  - Similarity between superpixels
  - Limbs

Superpixels labels

$y_i \in \{1, \ldots, C\}$

Limb segment labels

$l_p \in \{1, \ldots, C\}$

Unary potentials

- Simple features
- Person Mask
- Clothelets
- Shape features

Pairwise potentials

- Similarity between superpixels
- Limbs

Reduce complexity

More discriminative

Superpixels labels

$y_i \in \{1, \ldots, C\}$

Limb segment labels

$l_p \in \{1, \ldots, C\}$

Unary potentials

- Simple features
- Person Mask
- Clothelets
- Shape features

Pairwise potentials

- Similarity between superpixels
- Limbs

2D Pose Detector (Yang and Ramanan, CVPR 2011)

Problem specific

Superpixels labels
$y_i \in \{1, \dots, C\}$

Limb segment labels
$l_p \in \{1, \dots, C\}$

Unary potentials

- Simple features
- Person Mask
- Clothelets
- Shape features

Pairwise potentials

- Similarity between superpixels
- Limbs

Color, texture, and location histograms (Yamaguchi et al., CVPR 2012)

Input	GT	Unary

Superpixels labels

$y_i \in \{1, \ldots, C\}$

Limb segment labels

$l_p \in \{1, \ldots, C\}$

Unary potentials

- Simple features
- Person Mask
- Clothelets
- Shape features

Pairwise potentials

- Similarity between superpixels
- Limbs

Foreground/background segmentation by CPMC (Carreira and Sminchisescu, PAMI 2012)

Input GT Unary

Superpixels labels

$y_i \in \{1, \dots, C\}$

Limb segment labels

$l_p \in \{1, \dots, C\}$

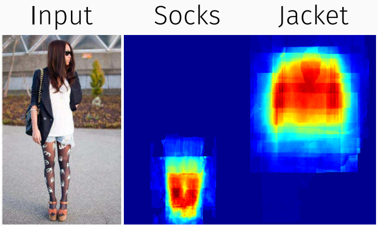
Unary potentials

- Simple features
- Person Mask
- Clothelets
- Shape features

Pairwise potentials

- Similarity between superpixels
- Limbs

Pose-conditioned garment likelihood



Superpixels labels

$y_i \in \{1, \ldots, C\}$

Limb segment labels

$l_p \in \{1, \ldots, C\}$

Unary potentials

- Simple features
- Person Mask
- Clothelets
- Shape features

Pairwise potentials

- Similarity between superpixels
- Limbs

Enriched SIFT descriptors with second order pooling (Carreira and Sminchisescu, ECCV 2012)

Input GT Unary

Institut de Robòtica i Informàtica Industrial

Superpixels labels

$y_i \in \{1, \ldots, C\}$

Limb segment labels

$l_p \in \{1, \ldots, C\}$

Unary potentials

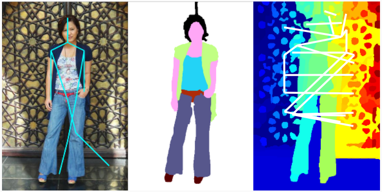
- Simple features
- Person Mask
- Clothelets
- Shape features

Pairwise potentials

- Similarity between superpixels
- Limbs

Shape, color, and texture similarity (Uijlings et al., 2013)

Input GT Pairwise



Superpixels labels

$y_i \in \{1, \ldots, C\}$

Limb segment labels

$l_p \in \{1, \ldots, C\}$

Unary potentials

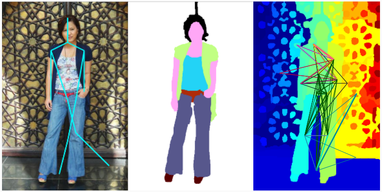
- Simple features
- Person Mask
- Clothelets
- Shape features

Pairwise potentials

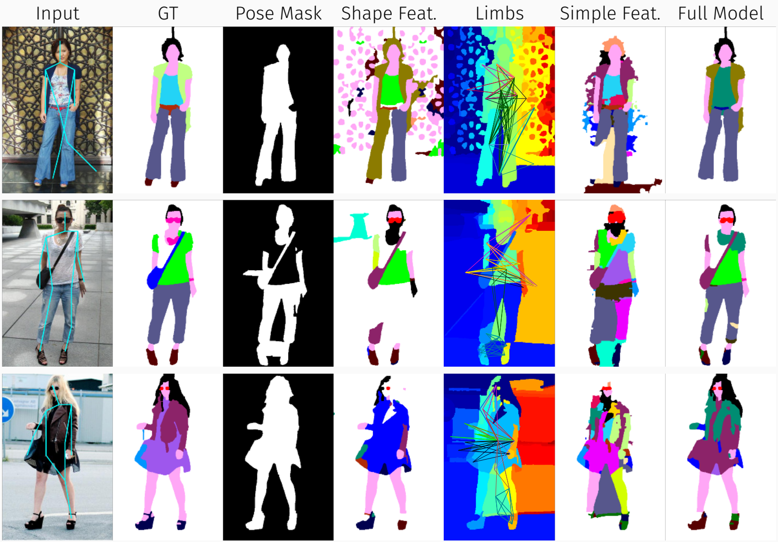
- Similarity between superpixels
- Limbs

Connect superpixels to limbs using 2D pose

Input GT Pairwise



FULL MODEL



Evaluation on the Fashionista dataset

- $\sim$700 images
- 29 and 56 class settings

Metric is Jaccard index or Intersection over Union: $\frac{t_p}{t_p+f_p+f_n}$

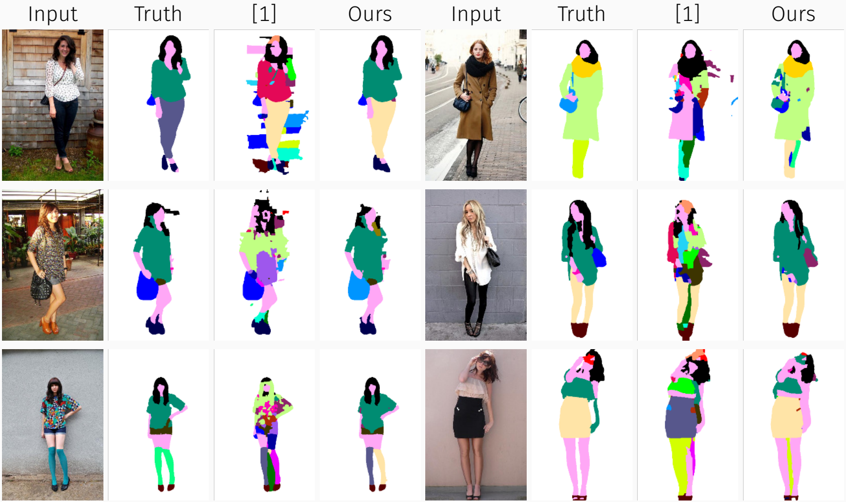
	29 Classes		56 Classes		
Method	[1]	Ours	[1]	[2]	Ours
Jaccard index	12.32	**20.52**	7.22	9.22	**12.28**

[1] K. Yamaguchi, H. Kiapour, L.E. Ortiz, T.L. Berg. Parsing clothing in fashion photographs. In CVPR, 2012.

[2] K. Yamaguchi, H. Kiapour, T.L. Berg. Paper Doll Parsing: Retrieving Similar Styles to Parse Clothing Items. In ICCV, 2013.

Institut de Robòtica i Informàtica Industrial

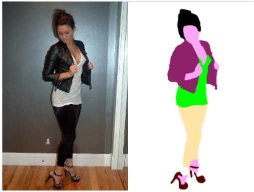
RESULTS - QUALITATIVE

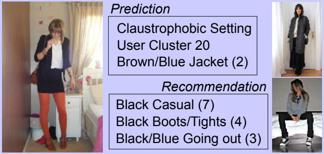


[1] K. Yamaguchi, H. Kiapour, L.E. Ortiz, T.L. Berg. Parsing clothing in fashion photographs. In CVPR, 2012.

Institut de Robòtica i Informàtica Industrial

Semantic Segmentation of Clothing





Modelling Fashionability

CONTRIBUTIONS

Large novel dataset (144,169 posts!)

Understand and model fashionability

- Give fashion advice!

LOS ANGELES, CA

466 FANS
288 VOTES
62 FAVOURITES

TAGS
CHIC
EVERDAY
FALL

COLOURS
WHITE-BOOTS

NOVEMBER 10, 2014

GARMENTS
White Cheap Monday Boots
Chilli Beans Sunglasses
Missguided Romper
Daniel Wellington Watch

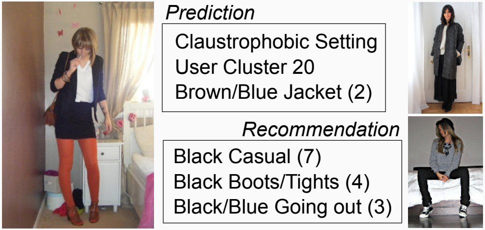
COMMENTS
Nice!!
Love the top!
cute
...

CONTRIBUTIONS

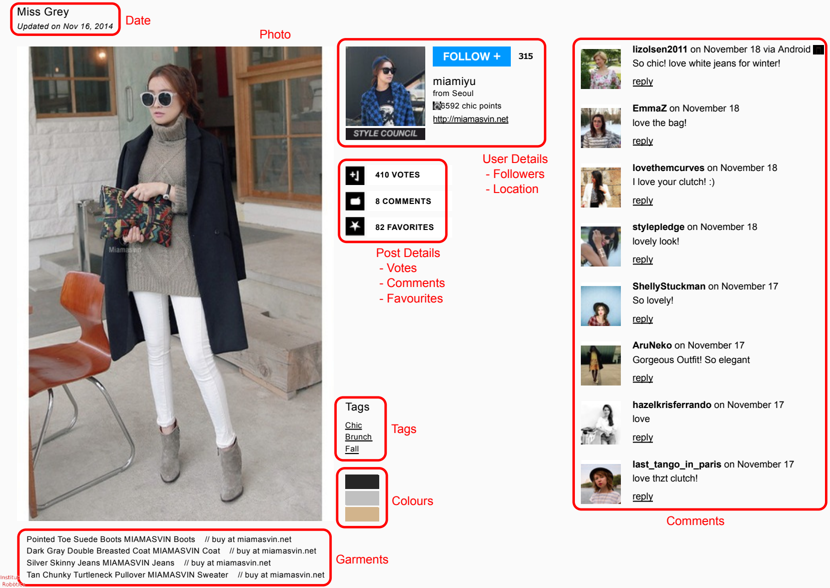
Large novel dataset (144,169 posts!)

Understand and model fashionability

- Give fashion advice!



DATASET


Miss Grey
Updated on Nov 16, 2014
Date
Photo
FOLLOW + 315
miamiyu
from Seoul
6592 chic points
http://miamasvin.net
STYLE COUNCIL
User Details
- Followers
- Location
410 VOTES
8 COMMENTS
82 FAVORITES
Post Details
- Votes
- Comments
- Favourites
Tags
Chic
Brunch
Fall
Tags
Colours
Pointed Toe Suede Boots MIAMASVIN Boots // buy at miamasvin.net
Dark Gray Double Breasted Coat MIAMASVIN Coat // buy at miamasvin.net
Silver Skinny Jeans MIAMASVIN Jeans // buy at miamasvin.net
Tan Chunky Turtleneck Pullover MIAMASVIN Sweater // buy at miamasvin.net
Garments
lizolsen2011 on November 18 via Android
So chic! love white jeans for winter!
reply
EmmaZ on November 18
love the bag!
reply
lovethemcurves on November 18
I love your clutch! :)
reply
stylepledge on November 18
lovely look!
reply
ShellyStuckman on November 17
So lovely!
reply
AruNeko on November 17
Gorgeous Outfit! So elegant
reply
hazelkrisferrando on November 17
love
reply
last_tango_in_paris on November 17
love thzt clutch!
reply
Comments

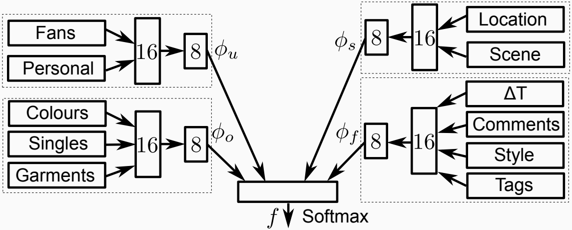
FEATURES

Feature	Dim.	Description
Fans	1	Number of user's fans.
ΔT	1	Time between post creation and download.
Comments	5	Sentiment analysis [1] of comments.
Location	266	Distance from location clusters.
Personal	21	Face recognition attributes.
Style	20	Style of the photography [2].
Scene	397	Output of scene classifier trained on [3].
Tags	209	Bag-of-words with post tags.
Colours	604	Bag-of-words with colour tags.
Singles	121	Bag-of-words with split colour tags.
Garments	1352	Bag-of-words with garment tags.

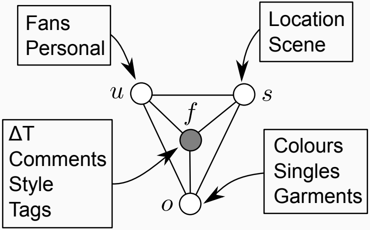
[1] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In EMNLP, 2013.
[2] S. Karayev, A. Hertzmann, H. Winnemoeller, A. Agarwala, T. Darrell. Recognizing image style. In BMVC, 2014.
[3] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In CVPR, 2010.

Model explicitly user (u), outfit (o), setting (s), and fashionability (f)

Features compressed using complementary deep networks



CRF Models Relationships between latent states



Model	Acc.	Pre.	Rec.	IOU	L_1
CRF	29.27	30.42	28.69	17.36	1.46
Deep Net	30.42	31.11	30.26	18.41	1.45
Log. Reg.	23.92	22.54	22.99	12.55	1.91
Baseline	16.28	-	10.00	1.63	2.32
Random	9.69	9.69	9.69	4.99	3.17

Recommendations: MAP estimate from conditional inference

Current Outfit:
Pink Outfit (3)

Recommendations:
Heels (8)
Pastel Shirts/Skirts (8)
Black/Gray Tights/Sweater (5)

Current Outfit:
Pink/Blue Shoes/Dress Shorts (3)

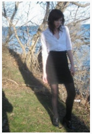
Recommendations:
Black/Gray Tights/Sweater (5)
Black Casual (5)
Black Boots/Tights (5)

Current Outfit:
Pink/Black Misc. (5)

Recommendations:
Pastel Dress (8)
Black/Blue Going out (8)
Black Casual (8)

Current Outfit:
Blue with Scarf (3)

Recommendations:
Heels (8)
Pastel Shirts/Skirts (8)
Black Casual (8)



Current Outfit:
Pink/Blue Shoes/Dress Shorts (3)

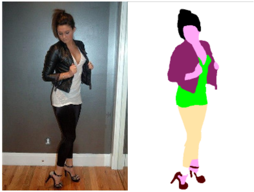
Recommendations:
Black Casual (7)
Black Heavy (3)
Navy and Bags (3)

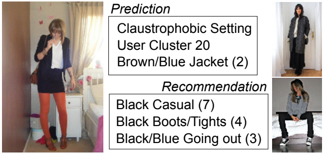
Current Outfit:
Formal Blue/Brown (5)

Recommendations:
Pastel Shirts/Skirts (9)
Black/Blue Going out (8)
Black Boots/Tights (8)

Institut de Robòtica i Informàtica Industrial

Semantic Segmentation of Clothing

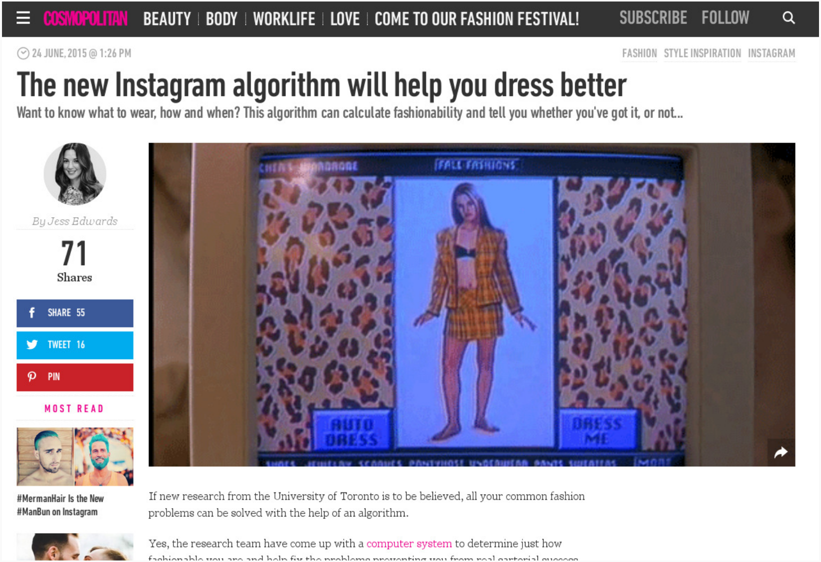




Modelling Fashionability

Institut de Robòtica i Informàtica Industrial

- Fashion is very challenging!
  - Hard for humans too!
- Proper framework is fundamental (CRF, . . .)
- Model must be tailored to the problem
- Potentials must be tailored to the problem
- Have to exploit as much information available
- Lots of room for improvement



COSMOPOLITAN BEAUTY BODY WORKLIFE LOVE COME TO OUR FASHION FESTIVAL! SUBSCRIBE FOLLOW

24 JUNE, 2015 @ 1:26 PM

FASHION STYLE INSPIRATION INSTAGRAM

The new Instagram algorithm will help you dress better

Want to know what to wear, how and when? This algorithm can calculate fashionability and tell you whether you've got it, or not...

By Jess Edwards

71 Shares

SHARE 55

TWEET 16

PIN

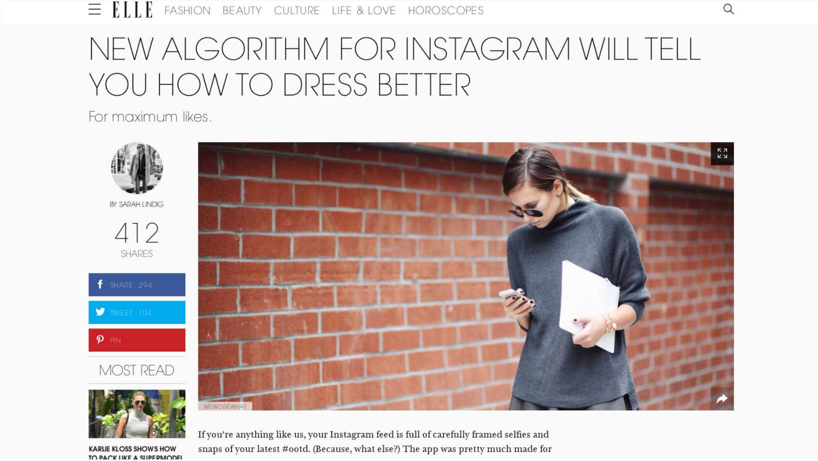
MOST READ

#MermanHair Is the New #ManBun on Instagram

If new research from the University of Toronto is to be believed, all your common fashion problems can be solved with the help of an algorithm.

Yes, the research team have come up with a computer system to determine just how fashionable you are and help fix the problems preventing you from real sartorial success

CONCLUSIONS



ELLE FASHION BEAUTY CULTURE LIFE & LOVE HOROSCOPES

NEW ALGORITHM FOR INSTAGRAM WILL TELL YOU HOW TO DRESS BETTER

For maximum likes.

BY SARAH LINDIG

412 SHARES

SHARE 294

TWEET 104

PIN

MOST READ

KARLIE KLOSS SHOWS HOW TO PACK LIKE A SUPERMODEL

WEWOREWHAT

If you're anything like us, your Instagram feed is full of carefully framed selfies and snaps of your latest #ootd. (Because, what else?) The app was pretty much made for

CONCLUSIONS

PORTADA MODA BELLEZA PASARELAS SUZY MENKES CELEBRITIES NOVIAS TV BLOGS FOROS

Inventan un software que mide tu nivel de estilo

Más sincero que la opinión de tu madre y de tu mejor amiga juntas, este programa puede valorar tu look, criticarlo y ofrecer consejos para mejorarlo

Beatriz Caballero — *@bethfevrier* — **Las ocasiones importantes, requieren *looks* importantes** y su aprobación pasa siempre por el vistode la amiga o persona de confianza de turno. Que levante la mano la que no se haya hecho una foto delante del espejo con el modelito escogido o en el probador de su tienda favorita y la haya enviado a ese grupo de *whatsapp*, esperando impacientemente el **veredicto**.

Pero este método, no nos engañemos, tiene fisuras... ¿Es esa amiga realmente sincera? ¿El gusto de tu amiga es el que más prevalece sobre tu propio criterio? ¿Ese grupo de *whastapp* es realmente *coo*l para valorar si ese es tu mejor *outfit*

Vogue Newsletter

Tu email

ok

He leído y acepto las Condiciones de uso así como la Política de protección de datos

Vogue *en* Facebook

A personas les gusta Vogue España

like

VOGUE LIVE

Institut de Robòtica i Informàtica Industrial

CONCLUSIONS

WIRED.CO.UK FASHION COMPUTER VISION MACHINE LEARNING TECHNOLOGY

THIS ALGORITHM JUST SOLVED FASHION

FASHION / 23 JUNE 15 / by KATIE COLLINS

318 shares
2 comments

Are you ever worried your friends aren't being honest with you about your taste in fashion? Well now you can turn to a computer in order to have your suspicions confirmed, or not, as the case may be. Don't be downcast if it turns out your outfits are not de rigeur, however -- the software can also help you plan a Cher from Clueless-style makeover.

Researchers at the University of Toronto have developed an algorithm that uses a combination of computer vision and

CONCLUSIONS

- Feature Point Descriptors
  - Deformation and Light Invariant (DaLI) Descriptor
  - Deep Convolutional Neural Network Descriptors
- Generative 3D Human Pose Models
  - Linear Latent Models
  - Directed Acyclic Graphs
  - Geodesic Finite Mixture Models
- 3D Human Pose Estimation
  - 3D Pose Estimation from Noisy Observations
  - Joint 2D and 3D Pose Estimation
- Fashion Understanding
  - Semantic Segmentation of Clothing
  - Modelling Fashionability

Simo-Serra, E., Ramisa, A., Alenyà, G., Torras, C., and Moreno-Noguer, F. Single Image 3D Human Pose Estimation from Noisy Observations. In IEEE Conference on Computer Vision and Pattern Recognition, 2012.

Simo-Serra, E., Quattoni, A., Torras, C., and Moreno-Noguer, F. A Joint Model for 2D and 3D Pose Estimation from a Single Image. In IEEE Conference on Computer Vision and Pattern Recognition, 2013.

Simo-Serra, E., Torras, C., and Moreno-Noguer, F. Geodesic Finite Mixture Models. In British Machine Vision Conference, 2014.

Simo-Serra, E., Fidler, S., Moreno-Noguer, F., and Urtasun, R. A High Performance CRF Model for Clothes Parsing. In Asian Conference on Computer Vision, 2014.

Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., and Moreno-Noguer, F. Fracking Deep Convolutional Image Descriptors. arXiv preprint arXiv:1412.6537, 2014.

Simo-Serra, E., Torras, C., and Moreno-Noguer, F. DaLI: Deformation and Light Invariant Descriptor. International Journal of Computer Vision, 2015.

Simo-Serra, E., Torras, C., and Moreno-Noguer, F. Lie Algebra-Based Kinematic Prior for 3D Human Pose Tracking. In International Conference on Machine Vision and Applications [best paper], 2015.

Simo-Serra, E., Fidler, S., Moreno-Noguer, F., and Urtasun, R. Neuroaesthetics in Fashion: Modeling the Perception of Fashionability. In IEEE Conference on Computer Vision and Pattern Recognition, 2015.

- Low to high level overview of human-centric computer vision
- Computer Vision and Machine Learning are highly intertwined
- Must stay up to date and exploit existing tools
- Code is available for most projects [1]

[1] http://www.iri.upc.edu/people/esimo/

Holistic models - multiple tasks at once

More real world applications

More features from state of the art (deep networks)

FUTURE DIRECTIONS

Holistic models - multiple tasks at once

More real world applications

More features from state of the art (deep networks)

THANKS

Thanks to my directors, collaborators, friends, family, supporters, and most importantly...

THANKS

Thanks to my directors, collaborators, friends, family, supporters, and most importantly...


Institut de Robòtica i Informàtica Industrial

QUESTIONS?

HTTP://WWW.IRI.UPC.EDU/PEOPLE/ESIMO/