TÉCNICO LISBOA

Institute for Systems and Robotics | LISBOA

Context-aware Person Re-identification via Fusion of Anthropometric and Gait Features

Athira Nambiar, Alexandre Bernardino,
Jacinto C. Nascimento and Ana Fred

BMVA technical meeting: Security and Surveillance,
26 April 2017, British Computer Society, London

LARSyS
Laboratory of Robotics and Engineering Systems

Computer and Robot Vision Lab

VIS LAB

Outline

1. Problem statement
    1.1 Motivation and scope
    1.2 Contributions
2. Anthropometric and Gait features
3. Proposed methodology
4. Experiments and Results
5. Conclusions

Surveillance

- Increased interest in surveillance technologies
- In the United Kingdom, there are between 4 million and 6 million CCTV surveillance [British Security Industry Association (BSIA)]
  - one for every eleven people
  - each Londoner is caught on camera **300 times** each day*

(*http://www.ibtimes.co.uk/britain-cctv-camera-surveillance-watch-london-big-312382)

Security & Surveillance

- Terror attacks boost calls for more surveillance

2013 bombing of the Boston marathon.

http://www.thejournal.ie/timeline-dzhokhar-tsarnaev-boston-bombing-2106664-May2015/

7 July 2005 London bombings

https://en.wikipedia.org/wiki/7_July_2005_London_bombings

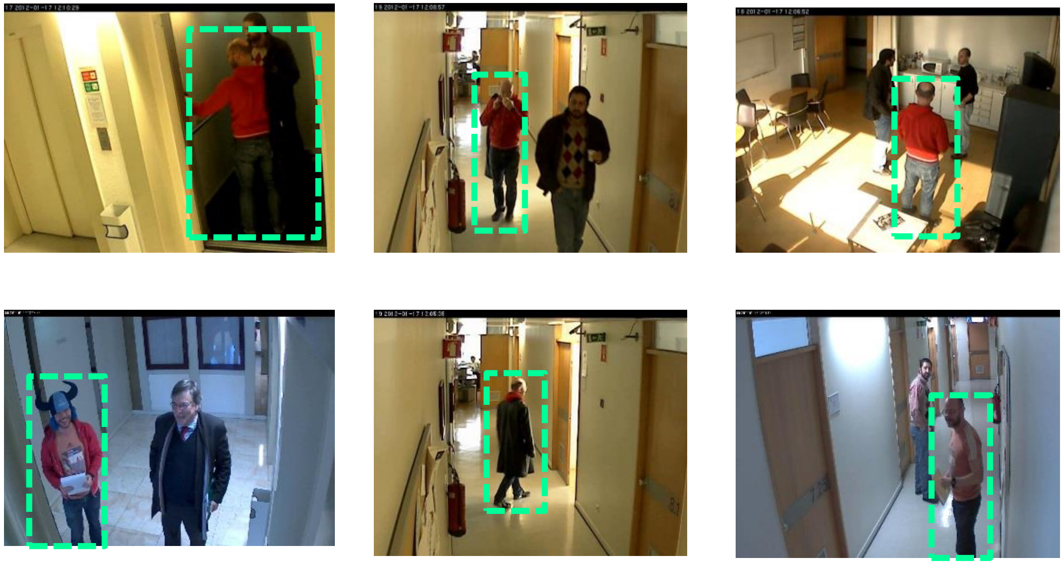
DUBAI ASSASSINATION: Hamas commander Mahmoud Al Mabhouh was killed in Dubai. (Getty Images)

http://www.arabianbusiness.com/for-hamas-murder-suspects-40450.html.

- Extensive research in the surveillance algorithms for automatic analysis of peoples identity/ behaviour.

Person Re-identification (Re-ID)

- Identify subject at different locations and different timings



*** HDA Person dataset: http://vislab.isr.ist.utl.pt/hda-dataset/ (IST-Lisboa)**

Challenges:

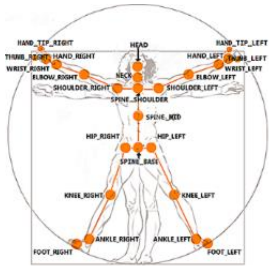
- ❑ varying background
- ❑ illumination changes
- ❑ Inter-camera variations
- ❑ occlusion
- ❑ view-point changes
- ❑ appearance changes over long term

Long-term view-point invariant person Re-ID

- Change in appearance over long periods of time

  - *Which features are robust to long term?*

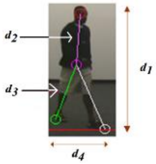
Anthropometry *Human gait*

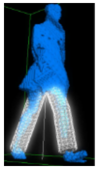


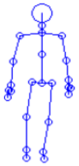
- Change in camera or subject pose

  - *How to get pose invariance?*

3D models

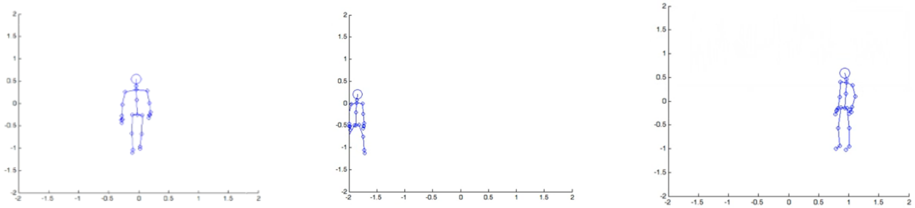






Key contributions

- **Pose-invariant database**, by collecting walking sequences in different directions using Kinect™ V.2 sensor



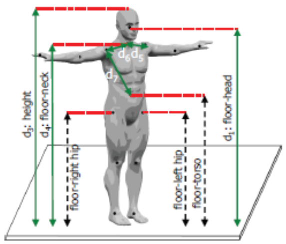
- Study of the **influence of various features** on Re-ID (individually and jointly) and impact of **Feature Selection**

- **'Context-aware ensemble fusion framework'** Re-ID system with **view-points as 'contexts'**
  " features depend strongly on the view-points"

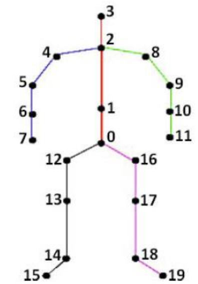
Related works

- Kinect based Re-ID

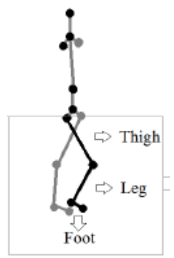
Pose dependent datasets !!



Barbosa et al. (2012)



Gianaria et al. (2014) ,



Andersson et al. (2015)

- Context

Context-based person identification framework for smart video surveillance

Liyan Zhang · Dmitri V. Kalashnikov · Sharad Mehrotra · Ronen Vaisenberg

Person re-identification with content and context re-ranking

Authors Authors and affiliations

Qingming Leng, Ruimin Hu, Chao Liang, Yimin Wang, Jun Chen

Person Re-Identification Ranking Optimisation by Discriminant Context Information Analysis

Jorge García[1], Niki Martinel[2], Christian Micheloni[2] and Alfredo Gardel[1]
[1]Department of Electronics, University of Alcala, Alcalá de Henares, Spain
[2]Department of Mathematics and Computer Science, University of Udine, Udine, Italy
jorge.garcia@depeca.uah.es, niki.martinel@uniud.it, christian.micheloni@uniud.it, alfredo@depeca.uah.es

1)I. B., Barbosa, M. Cristani, D.B. Alessio, L. Bazzani, and V. Murino (2012). Re-identification with RGB-D sensors. ECCV 2012.
2) E. Gianaria, M. Grangetto, M. Lucenteforte, and N. Balossino (2014). Human classification using gait features. Biometric Authentication.
3) V.O. Andersson, and R.M. Araujo(2015). Person identification using anthropometric and gait data from Kinect sensor. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence.

Outline

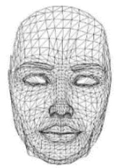
Biometrics for Re-ID

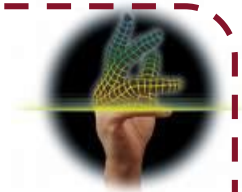
- Biometrics is the science of establishing the **identity of an individual** based on the physical, chemical or behavioral attributes of the person. - *Handbook of Biometrics*

HARD BIOMETRICS


Fingerprint


Iris / Retina

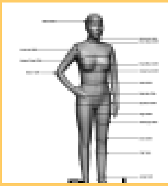

Face


Hand geometry

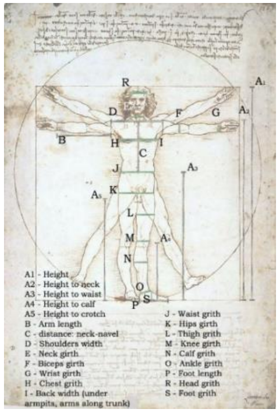
SOFT BIOMETRICS


Gender/ ethnicity


Hair style/ color

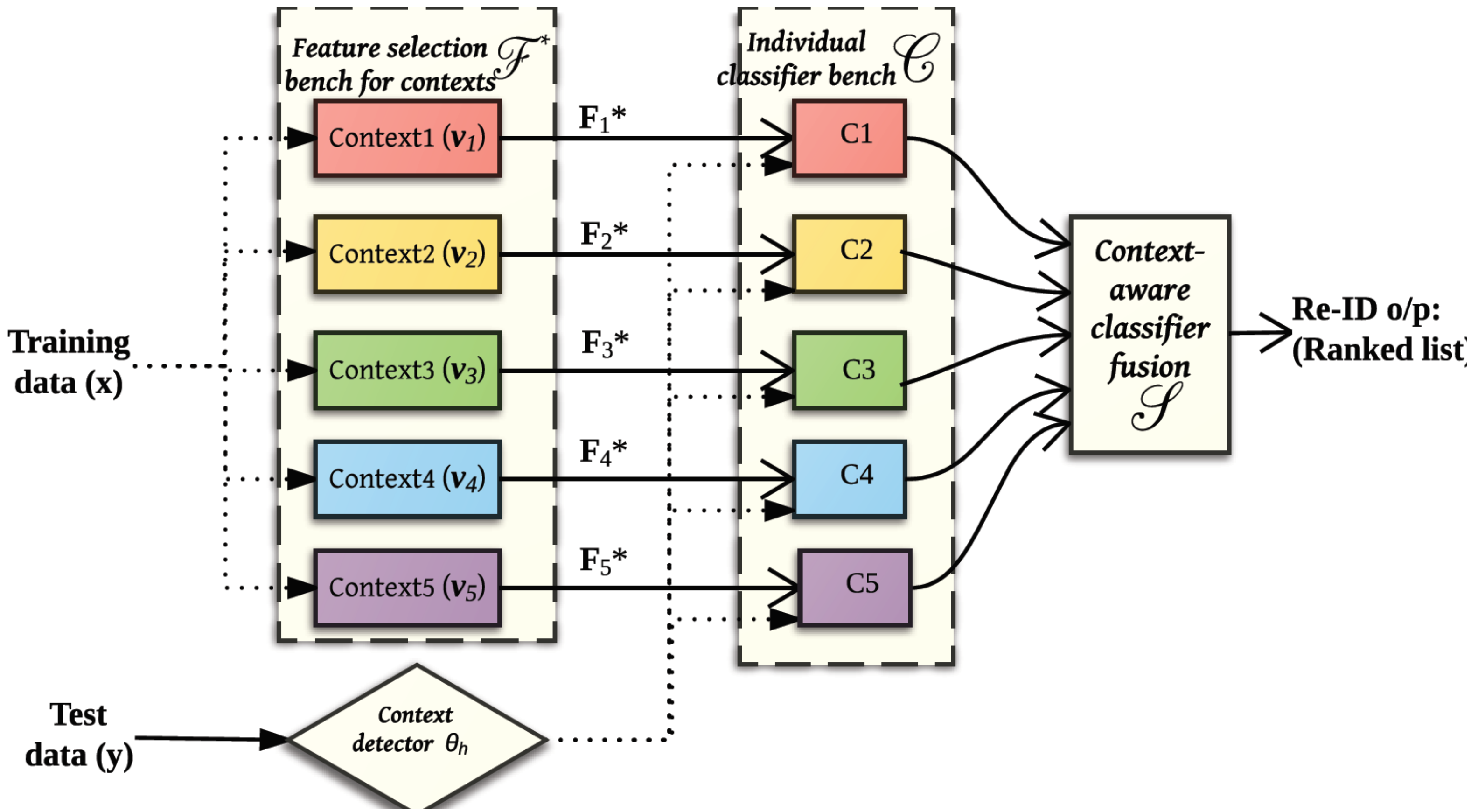

Body measurements


Gait



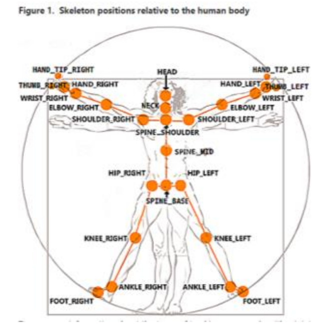
Outline

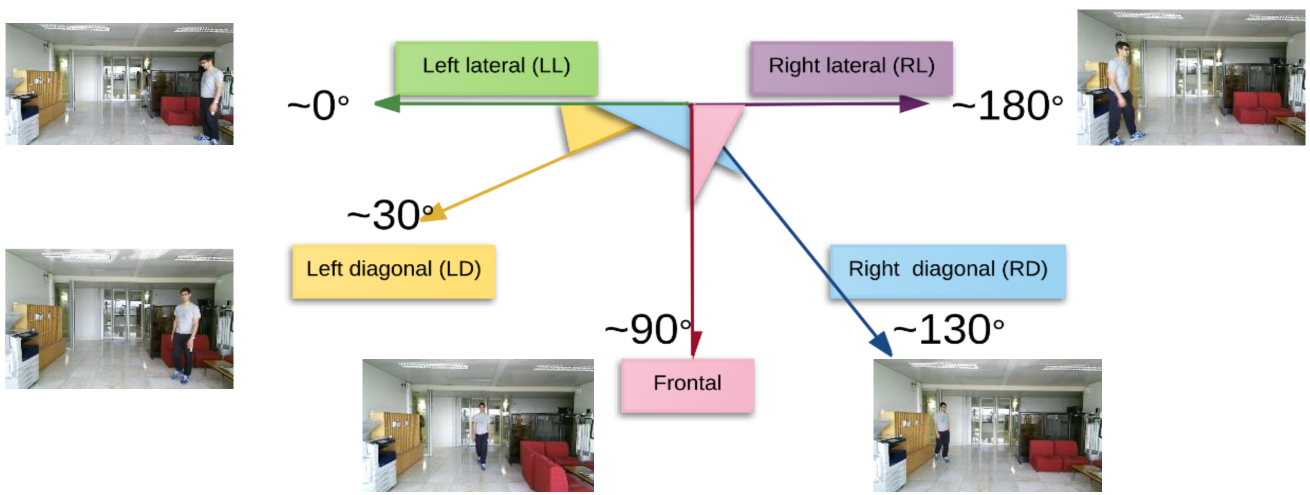
Context-aware view invariant Re-ID

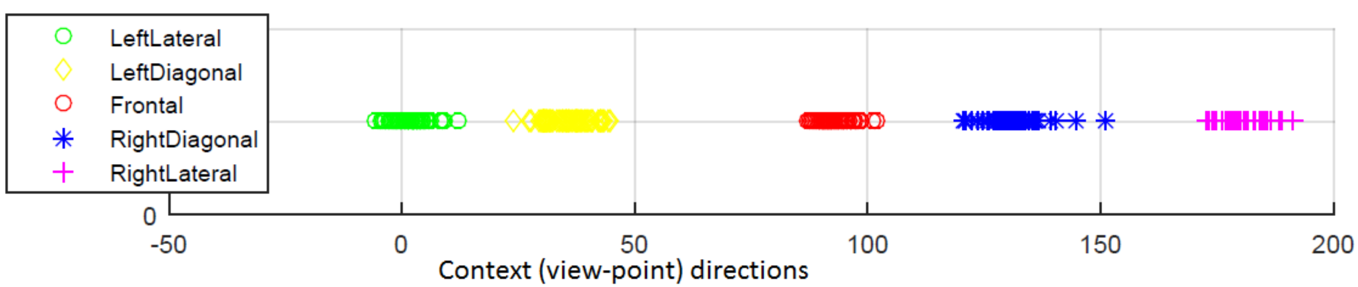


A. Database

- A new dataset with 20 people walking in 5 different directions acquired from Kinect v.2 (300 samples), suitable for pose-invariant Re-ID.







B. Feature extraction

Training data

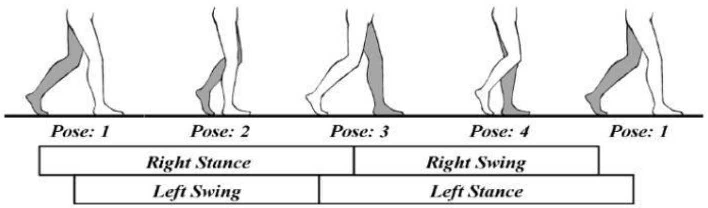
Gait cycle computation

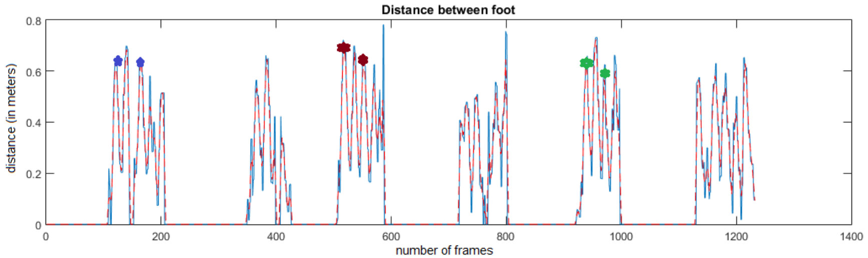
Feature extraction

Anthropometric Measurements (μ statistics)

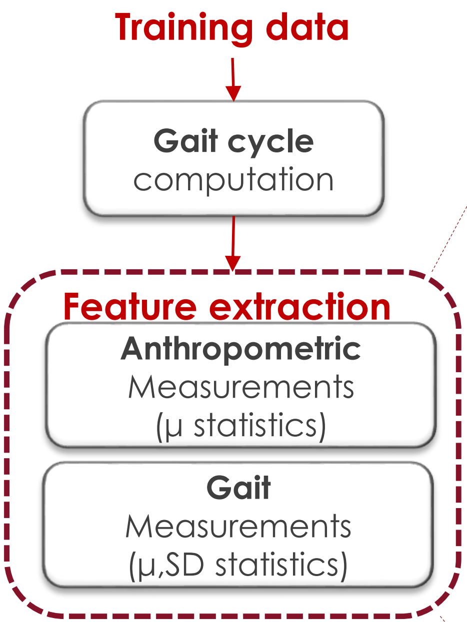
Gait Measurements (μ,SD statistics)

- the functional unit of gait
- the period of contact with the floor of the same foot





B. Feature extraction



Anthropometric features	Gait features	
Height-(1)	Hip angle(L&R)-(4)	Hip position(L&R)(x& y)-(8)
Arm length-(1)	Knee angle(L& R)-(4)	Knee position(L&R)(x& y)-(8)
Upper torso-(1)	Foot distance-(2)	Ankle position(L&R)(x& y)-(8)
Lower torso-(1)	Knee distance-(2)	Hand position(L&R)(x& y)-(8)
Upper-lower ratio-(1)	Hand distance-(2)	Shoulder position(L&R)(x& y)-(8)
Chestsize-(1)	Elbow distance-(2)	Stride-(1)
Hipsize-(1)	Head position(x& y)-(4)	Stride length-(1)
	Spine position(x& y)-(4)	Speed-(1)

- 7 anthropometric features (mean over a gait cycle)

(i.e., the static physical features defining the body measurements)

- 67 gait features (mean & standard deviation, over a gait cycle)

(i.e., dynamic features defining the kinematics in walking.)

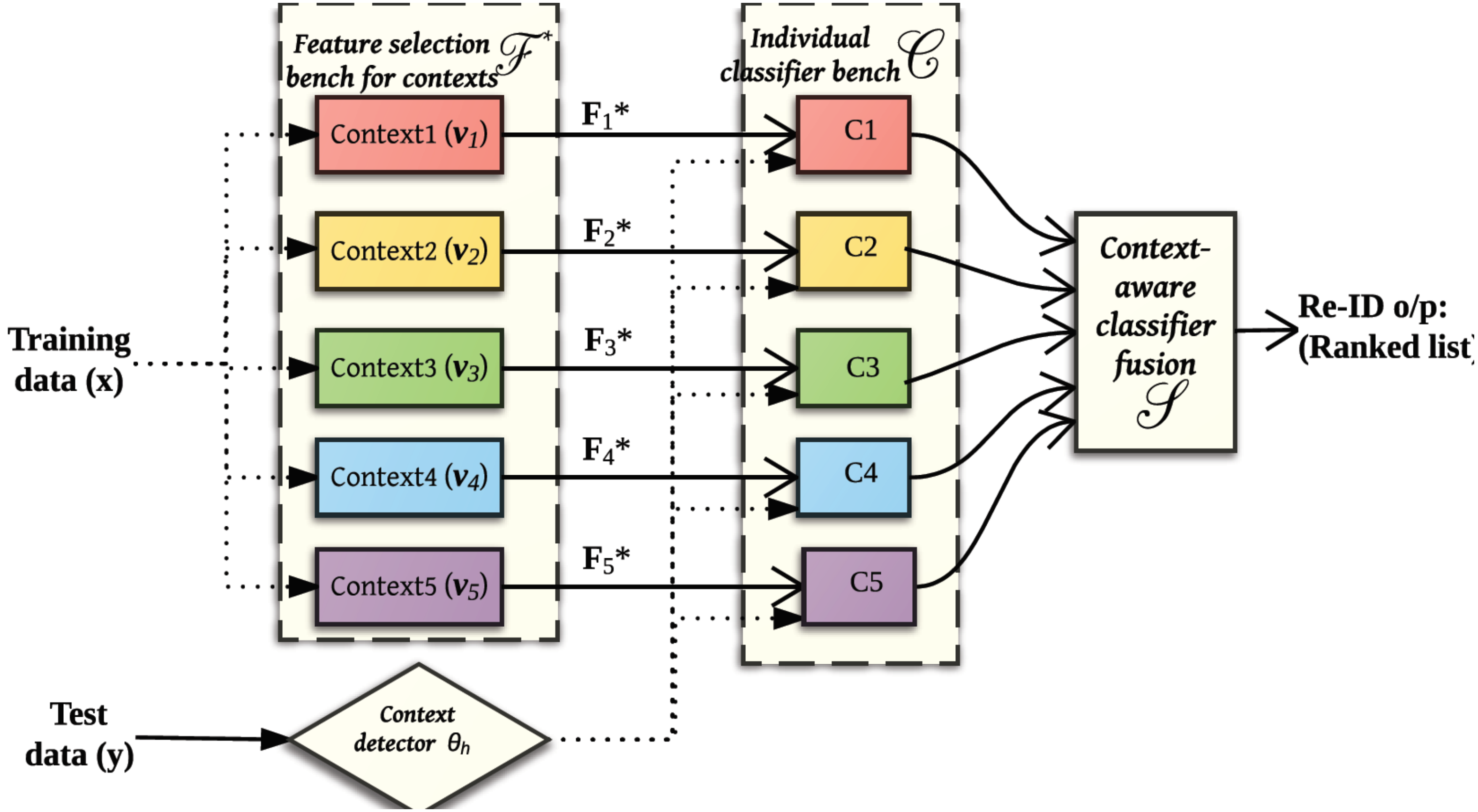
Feature selection -SFS

- Analyse the data in each context individually by leveraging a Feature Selection (FS) scheme in order to retain only the most discriminative and relevant features

- Sequential Forward Selection(SFS) algorithm

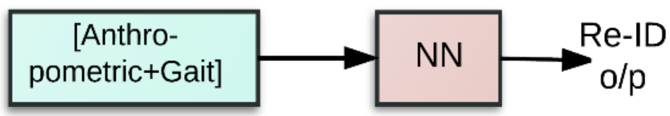
1. Start with the empty set $Y_0=\{\varnothing\}$
2. Select the next best feature $x^+ = \underset{x \notin Y_k}{\operatorname{argmax}}\left[J(Y_k + x)\right]$
3. Update $Y_{k+1}=Y_k+x^+$; $k=k+1$
4. Go to 2

Context-aware view invariant Re-ID

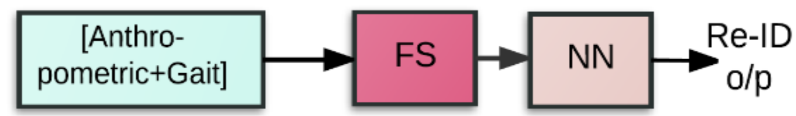


Feature selection and Fusion

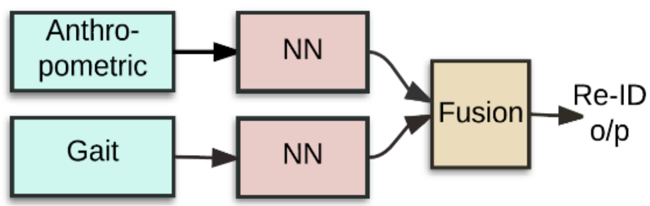
- Various Fusion-Feature selection schemes in order to combine anthropometric and gait features
- The best among the group and thus is considered as the 'de-facto' in our context-aware ensemble fusion framework, at the individual classifier bench.



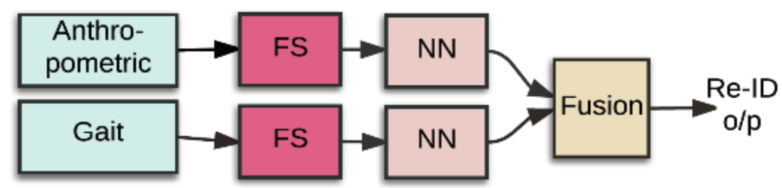
FL/NFS



FL/FS



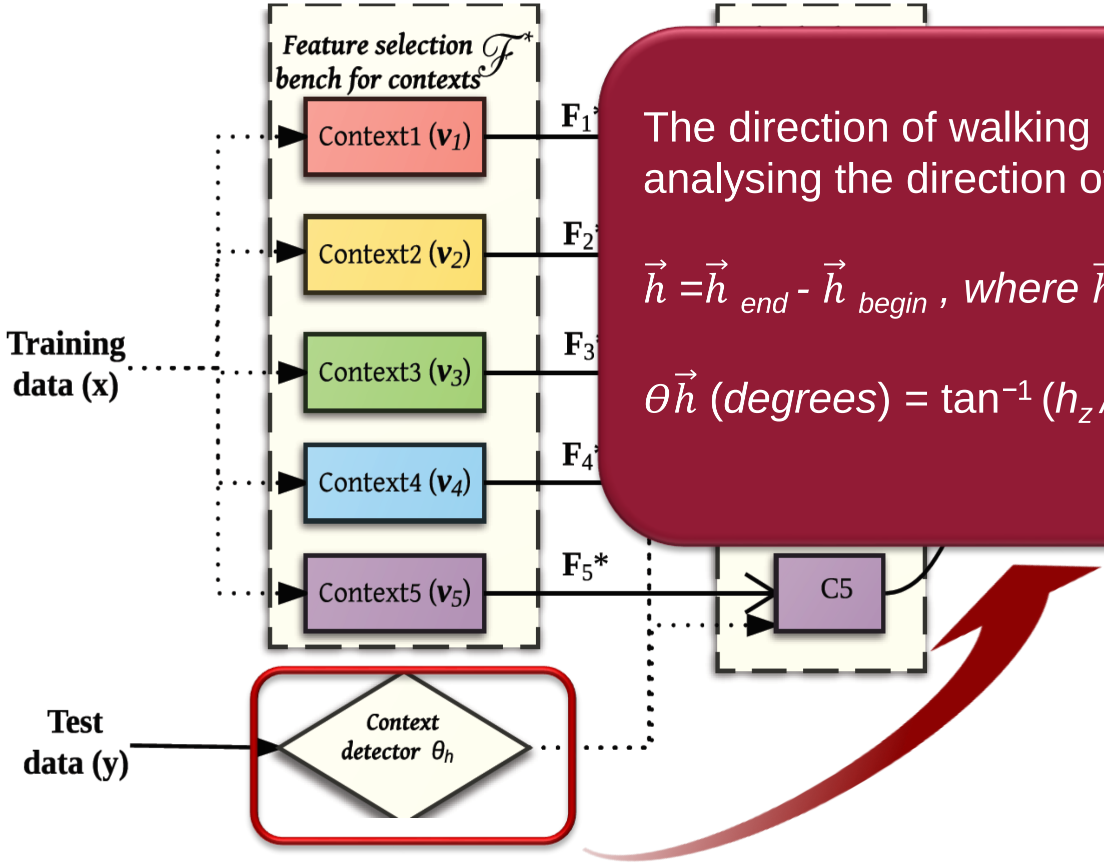
SL/NFS



SL/FS

FL- Feature level fusion; SL- Score level fusion
FS- Feature selection; NFS- No Feature Selection

C. Context-aware ensemble fusion



The direction of walking is estimated by analysing the direction of joint vector

$\vec{h} = \vec{h}_{end} - \vec{h}_{begin}$, where $\vec{h} = < h_x, h_y, h_z >$

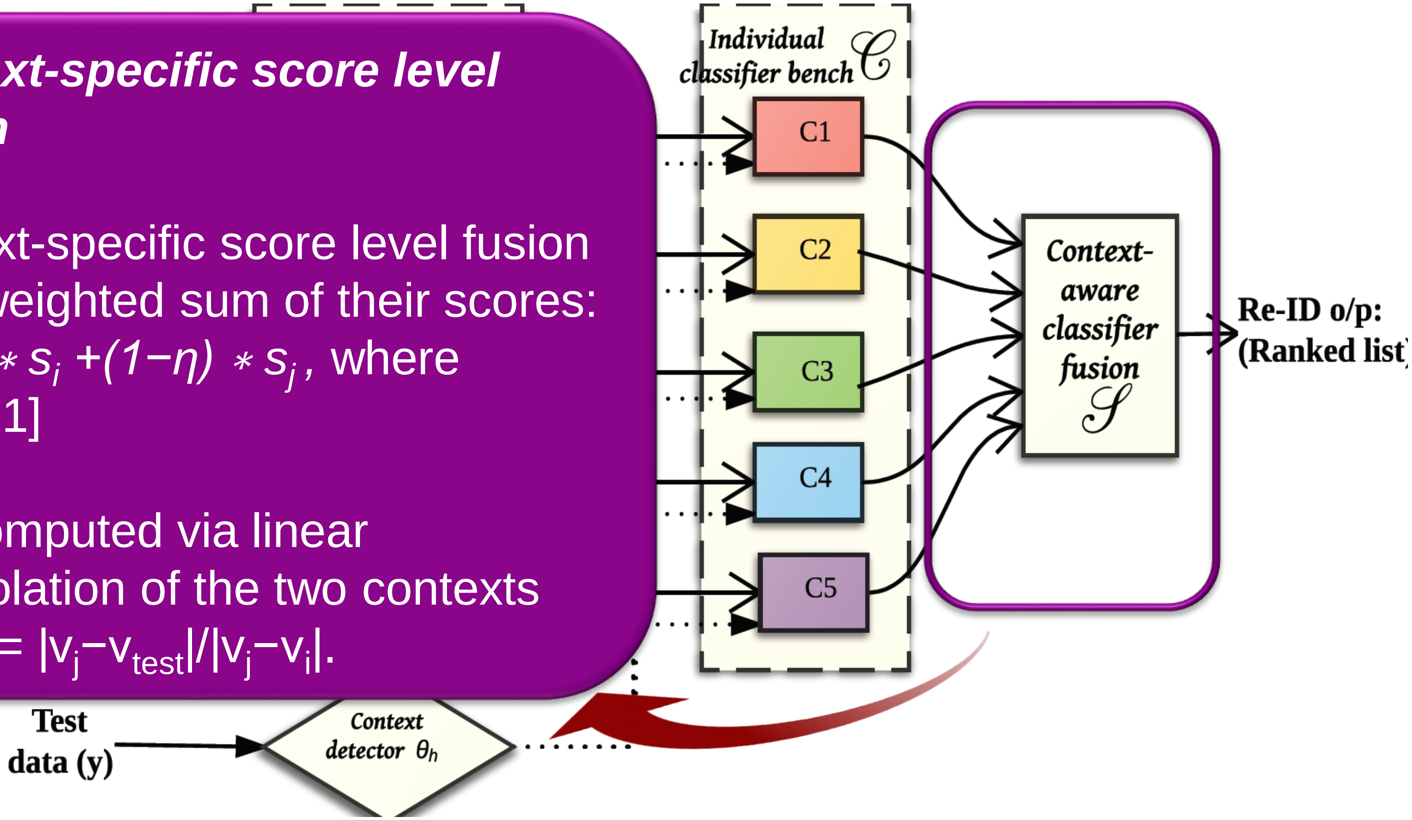
$\theta\vec{h}$ (degrees) $= \tan^{-1}(h_z/h_x) * 180/\pi$

C. Context-aware ensemble fusion

Context-specific score level fusion

Context-specific score level fusion S as weighted sum of their scores: $S = \eta * s_i + (1-\eta) * s_j$, where $\eta \in [0,1]$

η is computed via linear interpolation of the two contexts *i.e.*, $\eta = |v_j - v_{test}|/|v_j - v_i|$.

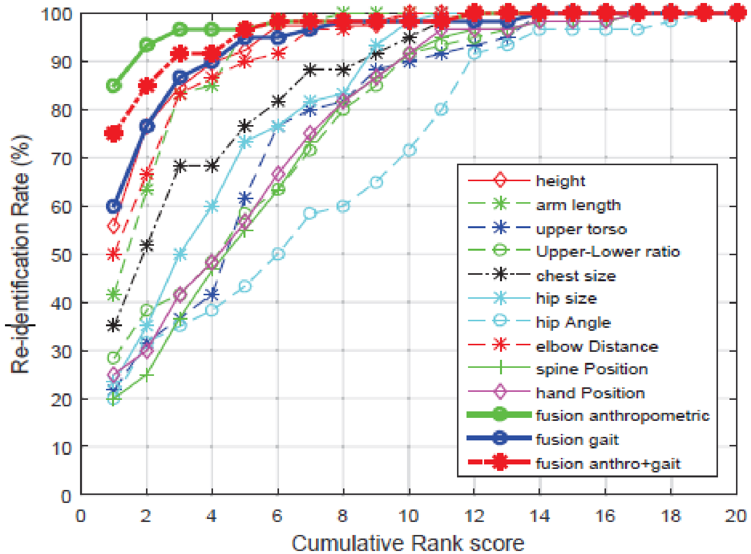


Outline

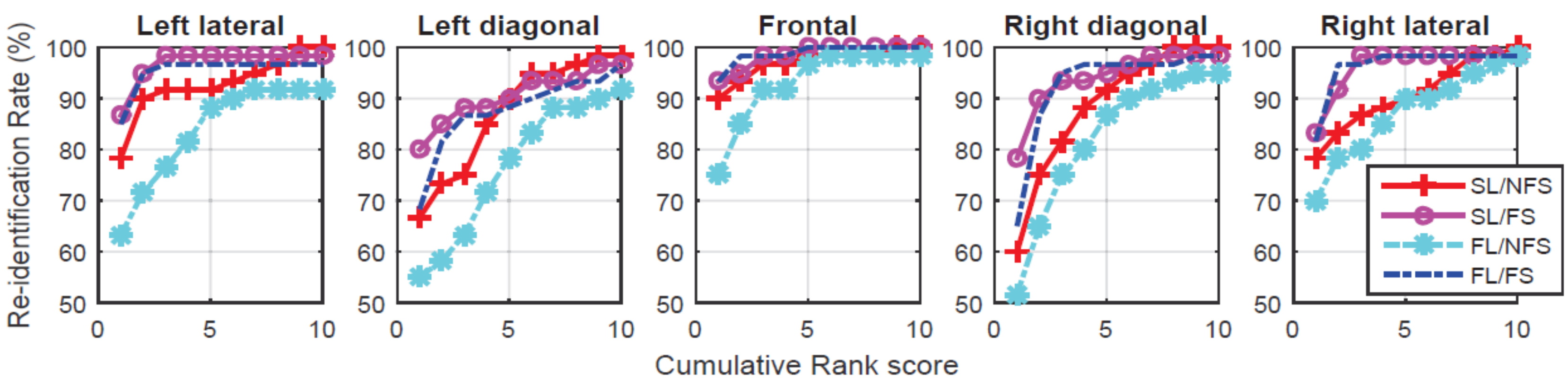
1. Problem statement
    - 1.1 Motivation and scope
    - 1.2 Contributions
2. Anthropometric and Gait features
3. Proposed methodology
4. Experiments and Results
5. Conclusions

A. Training the individual context-specific classifiers

Re-ID
performances of
individual as well as
fused features
in frontal context



-Various fusion-FS schemes for performance assessment

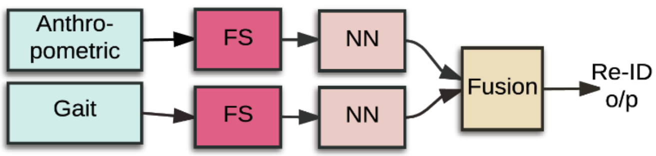


A. Training the individual context-specific classifiers

- *Feature selection (FS) improves Re-ID accuracy, compared to without FS (NFS).*
- *Score-level fusion works better than the feature level fusion in Re-ID.*
- *Overall performance of SL/FS is the best among the group and thus is considered as the 'de-facto' in our context-aware enser*

CONTEXT-SPECIFIC FEATURES SELECTED VIA SL/FS SCHEME, DURING THE TRAINING OF INDIVIDUAL CONTEXT CLASSIFIERS. ONLY 28 FEATURE SUBSET OUT OF WHOLE 74 FEATURES WERE SELECTED.

Feature	LL	LD	F	RD	RL	Feature	LL	LD	F	RD	RL
height	✓	✓	✓	✓	✓	spineY_{μ}	✓				
arm	✓	✓	✓	✓		lhipY_{μ}	✓				
upper	✓			✓		lkneeY_{μ}	✓	✓		✓	✓
lower		✓		✓	✓	rkneeY_{μ}	✓	✓		✓	
ULratio		✓		✓		rankleY_{μ}				✓	
chestsize		✓	✓	✓	✓	lhandX_{μ}			✓		
hipsize	✓		✓	✓		lhandY_{μ}	✓				
hipAngle			✓			$\text{lhandY}_{\text{SD}}$				✓	
$\text{kneeDist}_{\mu,\text{SD}}$			✓			rhandY_{μ}				✓	✓
$\text{handDist}_{\mu,\text{SD}}$			✓			lshouldY_{μ}	✓				
elbowDist_{μ}		✓	✓	✓		$\text{lshouldY}_{\text{SD}}$		✓			
$\text{elbowDist}_{\text{SD}}$				✓		rshouldY_{μ}					✓
headY_{μ}	✓	✓	✓		✓	$\text{rshouldY}_{\text{SD}}$			✓		
headY_{SD}			✓			strideLength	✓				✓



*Score Level Fusion with Feature Selection (**SL/FS**)*

B. Context-Specific Score Level Fusion

Context-aware:

(i) Using *single* context *(binary weighted)*

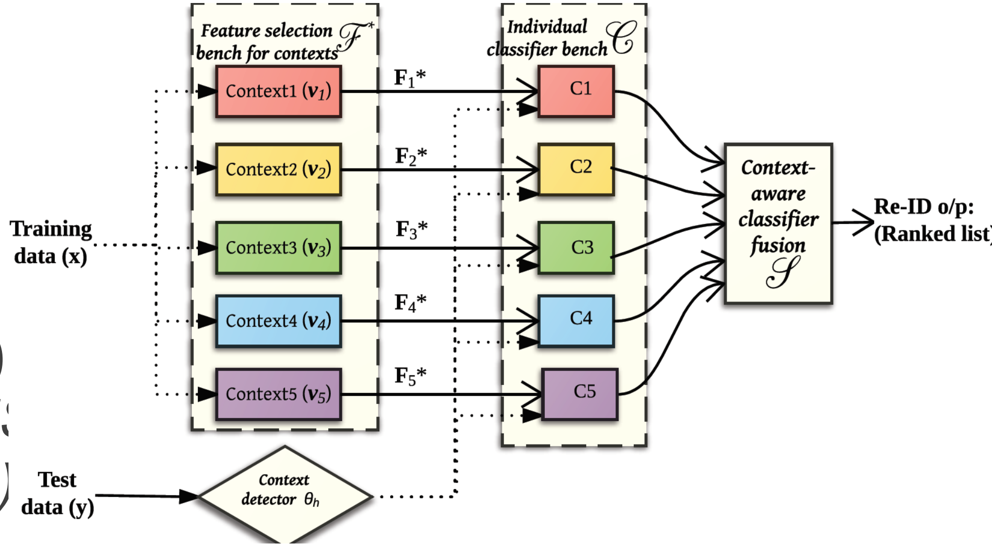
(ii) Using *two* contexts *(linear interpolated weights)*

Context-unaware (baseline)

(i) 'Pseudo baseline' (local F

(ii) 'Pure baseline' (global FS)

(iii) equal weights (0.2)



B. Context-Specific Score Level Fusion

	Context-unaware			Context-aware	
	No context (Pseudo baseline)	**No context (Pure baseline)**	**All contexts (equal weights)**	**1 context (binary weights)**	**2 contexts (adaptive weights)**
Anthropometric	25.33%	60.33%	45.67%	68.67%	68.00%
Gait Re-ID	26.67%	70.33%	53.33%	84.67%	85.67%
Overall Re-ID	74.33%	79.33%	71.33%	88.67%	88.33%
Processing time	25.7sec.	21.64sec.	25.92sec.	5.59sec.	10.47sec.

Fig: Results of classifier fusion showing our proposed context-aware classifier fusion against context-unaware baseline case studies. In the former cases, context detector module is enabled whereas in the latter cases, context-detector module is disabled. The experimental results showed that comparing to the Context-unaware systems, context-aware systems performed significantly faster (up to 4.5 times) and accurate (up to 17 percentage point better).

Outline

Conclusions

Top contributions

- A **long-term Re-ID** system leveraging anthropometric, gait and contexts.
- A novel **context-aware ensemble fusion framework** has been proposed towards long term Re-ID.
- Novel **Kinect based Re-ID dataset** with multiple view-points

Take home messages

- **Feature selection** always helps!
- **Score level fusion** outperforms Feature level fusion
- Comparing to the Context-unaware systems, **Context-aware systems** performed significantly **faster** (up to 4.5 times) and **accurate** (up to 17 percentage point better).

- Learning the contexts
- Multiple contexts (distance, people co-occurances etc.)
- Collecting more data in more random directions

Reference papers

1) **Context-Aware Person Re-identification in the Wild via fusion of Gait and Anthropometric features**, *A. Nambiar, A. Bernardino, J. Nascimento and A. Fred*, B-WILD Workshop at 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG). Washington DC, USA, 30 May - 3 June 2017

2) **Towards view-point invariant Person Re-identification via fusion of Anthropometric and Gait Features from Kinect measurements**, *A. Nambiar, A. Bernardino, J. Nascimento, A. Fred*, International Conference on Computer Vision Theory and Applications (VISAPP), Porto, Portugal, Feb. 2017

Computer and Robot Vision Laboratory (**VisLab**)

http://vislab.isr.tecnico.ulisboa.pt/

anambiar@isr.tecnico.ulisboa.pt